THE UNIVERSITY
of EDINBURGH

Investigating Agricultural and Biomedical Applications of Genome Editors in Large Animals

Rachel Huddart

Thesis presented for the degree of Doctor of Philosophy

University of Edinburgh, July 2015

Contents

List of Figures

List of Tables

Declaration

This thesis and the work presented within it are, unless otherwise stated, my own work. The work presented here has not been submitted for any other degree or professional qualification.

Rachel Huddart, July 2015

Acknowledgements

This PhD has not been the smoothest of rides and it's fairly safe to say that this thesis wouldn't exist with the help and support of a lot of people.

I have been very lucky to have ended up with two great supervisors who have been in my corner since the beginning of this project. Bruce, despite everything, has been a constant source of guidance, ideas and much-needed optimism and Simon has skilfully combined his frequent (and mostly successful) attempts to wind me up with an amazing amount of patience and calm when faced with my assorted panics and stupid questions. I really appreciate all the effort you two have put into getting me here, thank you. I'm also grateful to my thesis committee, Cheryl and Bernadette, for all their help during the project and especially to Bernadette for all her hard work and advice when things got tough.

In the lab, Claire has been a excellent molecular biology guru and a very cool lab aunt while Chris has been a massive help with everything from tissue culture to providing weekend cake. I also have to thank Maeve and Alex for their outstanding minion work and Spring and Sarah for being great fun to work with.

Outside of my group, there have been many other people who've helped me out along the way. I'd like to especially thank Lorna, Louise and Kirsty for keeping me laughing and Laura, Joni, Chloe, Catriona and Susan for the many drinks and pot luck dinners.

As always, my family have cheered me on every step of the way as well as reminding me that there is life outside of academia. Thank you and I promise I will stop studying now and get a proper job.

Finally, to Fabi for all the encouragement, support, jokes (of varying quality) and cups of tea which have kept me going over the last four years and for not complaining too much when you lost out to the lab at weekends. Danke.

Abstract

Large animal species, such as cattle, sheep and pigs, have great potential value to scientific research. This is due to their physiological similarity to humans, meaning they make excellent disease models in addition to their inherent agricultural value. However, the efficiency with which such animals can be created has been a critical barrier to their use in bioscience. Research into creating genetically modified large animals has not progressed as rapidly as research on smaller mammals, such as mice, for two main reasons. Firstly, technologies such as pluripotent stem cells, which are well established in rodents, are lacking for large animals. Secondly, large animals cannot produce as many offspring within a given time frame as mice or rats. This, combined with the low efficiencies and lack of precision of current transgenic methods, severely reduces the likelihood of obtaining an animal with a desired genotype within a viable amount of time.

Recently, new tools known as 'genome editors' have been developed to facilitate genetic modification of animals. The vastly enhanced efficiency of these editors in comparison to previous gene targeting methods, combined with the fact that genome editors do not require marker genes to be used, mean that creating genetically modified livestock is now far more feasible. This thesis investigates whether two types of genome editor, TALENs and CRISPR/Cas9, can be used to produce genetically modified large animals for a range of applications.

Genome editors were combined with interspecific blastocyst complementation techniques to produce chimeric rodents where the haematopoietic system is partially or fully derived from the donor cells. This work was carried out with a long-term aim of producing chimeric animals which could produce human organs suitable for transplantation. Initial blastocyst complementation experiments were carried out by injecting murine ESCs into wildtype rat blastocysts. One animal resulting from these injections showed chimerism in several tissues. Further experiments were carried out using rat ESCs and mouse blastocysts which were either Runx1$^{-/-}$

or Rag1$^{-/-}$, however no additional chimeras were identified. In addition to these experiments, TALENs and sgRNAs were designed against Runx1 and Rag1 in sheep and pigs in order to create a large animal model for future blastocyst complementation experiments.

Increasing animal productivity is a key step in meeting the demands of an increasing global population and tackling future food insecurities. TALENs and sgRNAs for use in the CRISPR/Cas9 system were created to target the myostatin gene in sheep. Myostatin is a negative regulator of muscle growth and animals which acquire natural inactivating mutations in both myostatin alleles exhibit a well-characterised double-muscled phenotype, where total muscle mass is about 20% greater than that of a wildtype animal. Embryo microinjections were carried out using both types of genome editor and two edited lambs were produced, one from each editor. The TALEN-edited lamb was mosaic for a deletion of arginine 283 which, upon further analysis of the muscle, did not appear to cause a significant phenotype. The CRISPR-edited lamb was heterozygous for a 20bp deletion, causing the formation of a premature stop codon and severe truncation of the mature myostatin protein. Based on data from other myostatin-knockout animals, including the Belgian Blue cattle breed, this truncated protein is not thought to be functional. To determine if this is indeed the case, the CRISPR-edited lamb is now part of a breeding programme to amplify the edited allele.

To discover if genome editors could be applied to create disease-resistant animals, the project focused on foot and mouth disease. Through a literature search and bioinformatic analysis of the bovine and porcine proteomes, three host genes which are cleaved by the virus were identified; eIF4A1, eIF4G1 and IKBKG. TALENs were designed to bind and cut at the FMDV protease cleavage sites in all three genes in order to disrupt protease cleavage and reduce viral replication by slowing viral disruption of the host translation and innate immune response pathways. Although none of the TALENs showed any signs of activity, this thesis sets out some potential directions for future work.

In conclusion, this thesis shows that, despite some technical issues, genome editors are a promising technology for the creation of genetically modified livestock.

Lay Summary

Genetic modification (GM) is the process by which the DNA of a living organism is altered to produce a desired characteristic, such as resistance to disease. This technique has many potential applications in different areas of science; from microbiology to plant science. However, established GM methods are notoriously inefficient and therefore can only really be applied to cases where enough offspring will be produced that there is a good chance of getting the desired modified organism.

Genome editors are a recently developed and easy-to-use GM method, with efficiencies far higher than those previously used. This means that animals which give birth to relatively few offspring, such as sheep and cattle, can now be genetically modified with a reasonable chance of success. This thesis explores several potential uses of GM livestock species. Firstly, we attempt to create animals which cannot produce blood, with the intention of "rescuing" these animals by injecting them with human blood cells and establishing them as a future source of human blood for transfusion. We also investigate the possibility of increasing the productivity of livestock, by increasing the amount of muscle on animals bred for meat and by making animals resistant to foot and mouth disease, which can be a major cause of financial loss for farmers.

The work in this thesis shows that it is possible to produce GM livestock animals with a relatively high efficiency. However, much of the research presented here ran into technical problems, demonstrating that there is plenty of scope for genome editors to be better understood and refined before they reach their maximum potential.

List of Abbreviations

ActRIIB	Activin receptor IIB
ALK	Activin receptor-like kinase
AML	Acute myeloid leukaemia
aNHEJ	Alternative non-homologous end joining
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATR	ATM + Rad-3 related
BEF	Bovine embryonic fibroblast
BHK	Baby hamster kidney
BLM	Bloom syndrome protein
BME	Basal Medium Eagle
BMP	Bone morphogenetic protein
BRCA1	Breast cancer 1, early onset
BRCA2	Breast cancer 2, early onset
BSA	Bovine serum albumin
Cas9	CRISPR associated protein 9
CBF	Core binding factor

CBFβ	Core binding factor, beta subunit
CD	Cluster of differentiation
CDK	Cyclin-dependent kinase
CEBPA	CCAAT/enhancer-binding protein alpha
CHD4	Chromodomain-helicase-DNA-binding protein 4
CHK	Checkpoint kinase
cNHEJ	Classical non-homologous end joining
CNS	Central nervous system
CRISPR	Clustered regularly interspaced short palindromic repeat
crRNA	CRISPR RNA
CSA	Cross-sectional area
CtIP	CTBP interacting protein
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
DNA-PKcs	DNA-dependent protein kinase, catalytic subunit
DSB	Double strand break
DTT	Dithiothreitol
EHT	Endothelial to haematopoietic transition
eIF4A1	Eukaryotic translation initiation factor 4 alpha 1
eIF4E	Eukaryotic translation initiation factor 4E

eIF4F	Eukaryotic translation initiation factor 4F
eIF4G1	Eukaryotic translation initiation factor 4 gamma 1
eIF4G2	Eukaryotic translation initiation factor 4 gamma 2
ERG	ETS-related gene
ESC	Embryonic stem cell
FCS	Foetal calf serum
FDA	Food and Drug Administration
FGF	Fibroblast growth factor
FLI1	Friend leukaemia integration 1 transcription factor
FLRG	Follistatin-related gene
FMD	Foot and mouth disease
FMDV	Foot and mouth disease virus
FOXO1	Forkhead box protein O1
GASP-1	GDF-associated serum protein 1
GASP-2	GDF-associated serum protein 2
GDF	Growth and differentiation factor
GEN1	GEN1 Holliday junction 5' flap endonuclease
GFP	Green fluorescent protein
GFI1	Growth factor independent 1 transcription repressor
GMEM	Glasgow modified Eagle's medium
GMO	Genetically modified organism
hCG	Human chorionic gonadotrophin

HDR	Homology-directed repair
HES1	Hairy and enhancer of split-1
HMGA2	High-mobility group AT-hook 2
HMGB1	High-mobility group protein B1
HSC	Haematopoietic stem cell
ICM	Inner cell mass
IFN	Interferon
IKBKG	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma
IKK	I kappa B kinase
iPS	Induced pluripotent stem
iPSC	Induced pluripotent stem cell
IRES	Internal ribosome entry site
IRF	Interferon regulatory factor
IVF	*In vitro* fertilisation
IVM	*In vitro* maturation
KMO	Kynurenine 3-monooxygenase
LB	Luria broth
LHE	LAGLIDADG homing endonuclease
LMO2	LIM domain only 2 (rhombotin-like 1)
L protease	Leader protease
LT-HSC	Long-term haematopoietic stem cell
LYL1	Lymphoblastic leukaemia associated haematopoiesis regulator 1

MAPK	Mitogen-activated protein kinase
MCR	Multigenic chain reaction
M-CSF	Macrophage colony-stimulating factor
MEF2C	Myocyte-specific enhancer factor 2C
MEM	Modified Eagle's medium
MgCl	Magnesium chloride
MHC	Myosin heavy chain
MKK6	MAPK kinase 6
MRE11	Meiotic recombination 11
mRNA	Messenger RNA
miRNA	Micro RNA
MSTN	Myostatin
MyoD	Myoblast determination protein
NADH	Reduced nicotinamide adenine dinucleotide
NBS1	Nibrin
NEAA	Non-essential amino acids
NF-χB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NHEJ	Non-homologous end joining
NK cell	Natural killer cell
NLS	Nuclear localisation signal
OEF	Ovine embryonic fibroblast
ORF	Open reading frame

PAM	Protospacer adjacent motif
PAXX	Paralog of XRCC4 and XLF
PBS	Phosphate buffered saline
PCGF1	Polycomb group RING finger protein 1
PCR	Polymerase chain reaction
PDX1	Pancreatic and duodenal homeobox 1
PHD	Plant homeodomain
PIKK	Phosphoinositol-3-kinase-like protein kinase
PMSG	Pregnant mare serum gonadotropin
PNK	Polynucleotide kinase
PNKP	Polynucleotide kinase/phosphatase
PRR	Pattern recognition receptor
RAD50	DNA repair protein Rad50 homolog
RAD51	DNA repair protein Rad51 homolog
RAD52	DNA repair protein Rad52 homolog
RNA	Ribonucleic acid
RNAi	RNA interference
RPA	Replication protein A
RRL	Rabbit reticulocyte lysate
RSS	Recombination signal sequence
RTEL	Regulator of telomere elongation helicase 1
RVD	Repeat variable diresidue

SALL1	Spalt-like transcription factor 1
SCID	Severe combined immunodeficiency
SCL	Stem cell leukaemia
SD	Sprague-Dawley
sgRNA	Short guide RNA/single guide RNA
shRNA	Short hairpin RNA
SLX4	SLX4 structure-specific endonuclease subunit
SUMO	Small ubiquitin-like modifier
SWI/SNF	Switch/Sucrose NonFermentable
TAK1	TGF-β-activated kinase 1
TALE	Transcription activator-like effector
TALEN	TALE-nuclease
TCR	T-cell receptor
TGF-β	Transforming growth factor beta
TOP3a	Topoisomerase (DNA) III alpha
tracrRNA	trans-activating CRISPR RNA
TRIM	Tripartite motif-containing
UTR	Untranslated region
WT	Wild-type
XLF	XRCC4-like factor
XRCC4	X-ray repair cross-complementing protein 4
ZFN	Zinc-finger nuclease

1. Introduction

1.1. Transgenic organisms

The ability to alter the genetic information contained within an organism has become a key part of scientific research in the last 50 years and is seen by some as part of the solution to global socio-economic issues, such as climate change and food security. The two key methods required to produced a transgenic organism are the ability to create a desired DNA molecule and the ability to insert this molecule into a cell or organism, both of which were first published in 1972. The bacterium *Escherichia coli* was the first organism to be successfully transformed with exogenous DNA (Cohen et al., 1972) and was soon followed by the report of the first recombinant DNA molecule, created by adjoining DNA molecules from the virus SV40 and the *E. coli* gal operon (Jackson et al., 1972). The production of recombinant DNA molecules improved with the introduction of restriction enzymes to construct plasmid molecules (Cohen et al., 1973), a technique which is still central to molecular cloning today.

Eukaryotic cells were transformed for the first time using the same principles behind those of bacterial transformation (Szybalska and Szybalski, 1962, Graham and van der Eb, 1973, Hinnen et al., 1978). The transformation of ribosomal RNA genes from *Xenopus* into *E. coli* showed that eukaryotic genes could be expressed in prokaryotic cells (Morrow et al., 1974), an observation that was later exploited to create bacteria which produced the regulatory hormone somatostatin (Itakura et al., 1977), leading to the use of transgenic organisms as bioreactors producing molecules of scientific and medical value.

The mouse was the one of the first multicellular organisms to be genetically modified by the injection of SV40 DNA into the blastocoel of mouse blastocysts (Jaenisch and Mintz, 1974). Analysis of animals resulting from the injections showed the presence of SV40 DNA in tissues of 34% of the adult mice analysed. Pronuclear injection allowed donor DNA to be

stably integrated into recipient genomes and passed down to descendants through the germline (Gordon and Ruddle, 1981). Since then, other multicellular organisms such as plants (Bevan et al., 1983), nematodes (Stinchcomb et al., 1985), insects (Rubin and Spradling, 1982), fish (Maclean and Talwar, 1984), birds (Bosselman et al., 1989) and large mammals such as pigs, sheep and primates (Hammer et al., 1985, Chan et al., 2001) have been successfully genetically modified by microinjection.

The isolation and culture of ES cells (Evans and Kaufman, 1981, Thomson et al., 1998, Buehr et al., 2008) and production of iPS cells (Takahashi and Yamanaka, 2006) has provided another route to producing genetically modified animals (Thomas and Capecchi, 1987, Doetschman et al., 1987, Tong et al., 2010, Okita et al., 2007). However, this resource is only available for a select number of species, such as mice and rats. Although iPS lines showed initial promise in providing suitable pluripotent cells for large animals transgenics, the lack of suitable culture conditions mean that this is not yet a reality. For species where stem cell lines are not available, cloning by SCNT provides another route to producing transgenic animals (Wilmut et al., 1997). Although the efficiency of producing live transgenic animals by cloning is low, and potentially prohibitive to laboratories which don't have appropriate resources, both SCNT and targeted stem cells have an advantage over embryo microinjection as the donor cells and modification can be characterised prior to their use.

Although new reagents and techniques such as viral vectors and transposons (Terwilliger et al., 1989, Sorge and Hughes, 1982, Berkner, 1988, Dupuy et al., 2002), Cre-lox recombination (Hamilton and Abremski, 1984), GFP and other marker genes (Chalfie et al., 1994) and miRNAs to induce gene knockdown (Fire et al., 1998) have helped to refine the process of genetic modification, making changes to endogenous genes and making subtle changes to DNA (i.e. mutating single nucleotides) is still a complex and inefficient process.

1.2. Genome editors

Genome editors are a recently developed technology which allow genetic modifications to be carried out with unprecedented accuracy and efficiency. Each genome editor has two main components; a customisable DNA-binding molecule, which is usually a protein or RNA molecule, and an endonuclease which relies on the DNA-binding molecule for its cleavage spe-

cificity (i.e. it has no target sequence requirements of its own). When introduced into a cell, genome editors induce double-strand breaks (DSBs) in the DNA. These breaks, in turn, stimulate the cell's DNA damage response and eventual repair of the DNA. Repair of DSBs in DNA tends to occur through one of two main mechanisms; non-homologous end joining (NHEJ) or homology-directed repair (HDR). From the perspective of using genome editors, NHEJ is the pathway exploited when the researcher wishes to knock out or disrupt a gene, as NHEJ's error-prone propensity makes it likely that inactivating mutations will be introduced into the DNA. When a more precise modification is required, a donor DNA template containing the required alteration can be introduced into the cell along with the genome editors and incorporated into the genome by HDR. In addition, both NHEJ and HDR can be used to mediate the deletion of DNA fragments. The DNA damage response and DNA repair by NHEJ or HDR are described in Section 1.2.1 below.

Genome editors are particularly useful in cases of gene targeting, where use of an endonuclease to introduce DSBs at the target site can increase the efficiency of targeting by 2 to 3 orders of magnitude (Rouet et al., 1994). More specifically, genome editors can increase the frequency of gene targeting from 1 in $1x10^6$-$1x10^7$ without an editor to 10% of the total cell population when an editor is used (Park and Telugu, 2013). This leap in efficiency also makes the production of cells where both alleles have been targeted more likely and, unlike other methods of genetic modification, there is no requirement to include marker genes, such as those encoding antibiotic resistance or fluorescent proteins.

1.2.1. DNA repair

1.2.1.1. DNA damage response

An outline of the DNA damage response is given below. For a detailed review, see Ciccia and Elledge (2010).

DNA damage, including DSBs caused by genome editors, is detected by DNA damage response proteins known collectively as phosphoinositol-3-kinase-like protein kinases or PIKKs. The individual PIKKs are the kinases DNA-PK (a protein complex of DNA-PKcs and Ku), ATM and ATR. It is known that ATM responds to DSBs where the broken ends are permanently blocked (i.e. they do not end with a 5´ phosphate or 3´ hydroxyl group), which accounts

for around 10-20% of the total DSBs experienced by a cell (Alvarez-Quilon et al., 2014). However, it is presently unclear which, if any, of the known PIKKs signal the presence of the remaining 80-90% of DSBs and by what mechanism this signalling occurs.

Once the presence of DNA damage is detected, ATM and DNA-PK are activated by the MRE11 complex, which consists of the nuclease MRE11, the DNA-binding protein RAD50 and the adaptor protein NBS1. The MRE11 complex is recruited to the site of the DSB (Lisby et al., 2004, Lee and Paull, 2005, 2004, Bakkenist and Kastan, 2003) and, in turn, recruits histone H2AX by phosphorylation (Furuta et al., 2003, Rogakou et al., 1998), which itself recruits more DNA repair factors (Celeste et al., 2002). ATM also prevents transcription from occurring near the site of the DSB (Shanbhag et al., 2010) and induces chromatin relaxation across the whole genome, preventing further DNA damage (Ziv et al., 2006). Activated ATM and ATR phosphorylate mediator proteins, such as CHK1 and CHK2. This has the effect of slowing or halting cell cycle progression and amplifies the DNA damage signal (Chehab et al., 2000, Liu et al., 2000, Matsuoka et al., 2000). This results in the activation of the ubiquitin-protein ligase BRCA1 (Cortez et al., 1999), which is involved in the regulation of the S-phase and G2-M checkpoints, as well as the stabilisation and activation of p53 by ATM and the helicase CHD4 (Kastan et al., 1992, Polo et al., 2010), which act at the G1-S cell cycle checkpoint. As well as cell cycle arrest, p53 can also initiate apoptosis or senescence in response to DNA damage.

Once cell cycle arrest has occurred, the cell then repairs the DNA by either NHEJ or HDR, described below.

1.2.1.2. Non-homologous end joining (NHEJ)

Two distinct NHEJ pathways have been identified; classical NHEJ (cNHEJ), the more studied of the two, and alternative NHEJ (aNHEJ). Both pathways are error-prone and are likely to introduce mutations into a repaired DNA site. They are generally exploited in genome editing techniques when an inactivating mutation is required, but the exact nature of the mutation is not important.

Classical NHEJ was first described in the 1980s following studies in eukaryotic cells (Wilson et al., 1982, Roth et al., 1989, 1985, Roth and Wilson, 1986) and is the major DNA repair pathway in eukaryotes, although it has a tendency to leave small indels in the repaired DNA.

cNHEJ is initiated when the Ku heterodimer, consisting of the proteins Ku70 and Ku80/Ku86, forms a ring around free DNA ends resulting from a DSB (Mari et al., 2006). Ku binds to the DNA sugar backbone, allowing it to bind DNA in a sequence-independent manner (Walker et al., 2001). The Ku heterodimer protects the DNA ends from further damage and provides a scaffold for other factors involved in cNHEJ, such as the scaffold protein XRCC4 (Mari et al., 2006) and DNA-PKcs (Uematsu et al., 2007). XRCC4 enters the nucleus and recruits DNA ligase IV to form the XRCC4-Ligase IV complex, which stimulates the ligase's activity (Grawunder et al., 1997).

DNA-PKcs stimulates synapsis of the broken DNA ends and recruits enzymes involved in end processing and ligation (DeFazio et al., 2002). Broken DNA ends which still have a 5' phosphate group and a 3' hydroxyl group can be directly ligated to each other at this point and the break repaired. DNA ends which have undergone greater levels of damage are first processed by enzymes such as Artemis, which opens DNA hairpins (Ma et al., 2002), the Ku heterodimer, which removes abasic nucleotides from the DNA (Strande et al., 2012) and PNKP, a phosphatase which adds and removes phosphate groups so that the DNA ends can be ligated together (Chappell et al., 2002). Several DNA polymerases are used to fill any gaps in the DNA strands at the site of a DSB, as reviewed in detail by Moon et al. (2007). This polymerase activity is mediated by the Ku-XRCC4-Ligase IV-XLF complex and PAXX, a scaffold protein related to XRCC4 (Akopiants et al., 2009, Ochi et al., 2015). DNA ends are finally ligated together by the XRCC4-Ligase IV complex (Grawunder et al., 1997, Ahnesorg et al., 2006), an action promoted by the DNA-binding protein XLF (Riballo et al., 2009). The XRCC4-Ligase IV complex can ligate across gaps (Gu et al., 2007) and can ligate each DNA strand individually (Ma et al., 2004). This allows one of the DNA strands to be repaired and protected if the other strand is not able to be ligated. The NHEJ complex then dissociates from the DNA via an unknown mechanism, although protein ubiquitination is thought to play a role (Davis and Chen, 2013).

Alternative NHEJ is considered to be more error-prone than cNHEJ and can cause larger mutations. It was first observed in cells which lacked crucial cNHEJ proteins, such as Ku80, (Liang and Jasin, 1996), XRCC4 (Kabotyanski et al., 1998), Ku70 (Boulton and Jackson, 1996) and Rag2 (Corneo et al., 2007). aNHEJ uses areas of microhomology to direct repair (Boulton and Jackson, 1996, Kabotyanski et al., 1998, Guirouilh-Barbat et al., 2007) and can often cause

chromosomal translocations due to unstable DNA ends (Simsek and Jasin, 2010, Zhu et al., 2002). The exact mechanism of aNHEJ has yet to be fully elucidated and it is possible that there may be multiple aNHEJ pathways (Simsek et al., 2011). However, there is evidence that RAD52, CtIP, the MRE11 complex and DNA ligase III may all have roles to play in this process (Boulton and Jackson, 1996, Zhang and Jasin, 2011, Lee-Theilen et al., 2011, Deriano et al., 2009, Xie et al., 2009, Simsek and Jasin, 2010). Although it was discovered in cNHEJ-incompetent cells, aNHEJ is also active in WT cells and accounts for about 10% of total NHEJ activity within a cell (Corneo et al., 2007).

1.2.1.3. Homology-directed repair (HDR)

Homology-directed repair is thought to only be active in the S and G2 phases of the cell cycle and, indeed, NHEJ is suppressed by Cdk1 at the G2 phase and by BRCA1 at S phase to allow HDR to occur (Zhang et al., 2009, Davis et al., 2015). Histone modifications and the position of the DSB within the nucleus also play a role in determining whether DNA will be repaired by NHEJ or HDR (Pai et al., 2014, Lemaitre et al., 2014). This could have implications on the success of genome editing, particularly in embryos, as if a cell is not in the correct cell cycle phase, HDR is far less likely to occur. Synchronising cell cycles and transfection of a donor template during a more HDR-permissive phase of the cell cycle has been shown to increase donor integration rates (Rivera-Torres et al., 2014).

DNA is prepared for HDR by the loading of the MRE11 complex onto the free DNA ends and subsequent activation of ATM (Lee and Paull, 2004). RAD50 is recruited to the MRE11 complex and may help to keep the DNA ends close to each other by forming a bridge between them (de Jager et al., 2001). ATM regulates resection of DNA, where small deletions are made at the 5' ends of DNA breaks. This causes Ku complexes to dissociate from the DNA and pushes cells to be more committed to HDR than NHEJ (Zhang et al., 2009). RPA complexes then bind to single-stranded 3' ends (Coverley et al., 1991, 1992), allowing assembly of RAD51 filaments by SUMOylation of RAD51 monomers (Golub et al., 1998, Shima et al., 2013), mediated by BRCA1 and BRCA2 (Esashi et al., 2005). RAD51 filaments are known to stimulate strand invasion into an homologous sequence or, in the case of genome editing, the externally supplied template containing the desired mutation. In prokaryotes, RecA (the RAD51 homolog) filaments also mediate the search for a homologous sequence in the genome, prior to strand

invasion (Lesterlin et al., 2014). Once it is paired with a homologous sequence, the invading strand is repaired by extension by polymerases and annealing to the other end of the break, which is regulated by the helicase RTEL (Barber et al., 2008). This leads to D-loop formation, which causes capture of the second end and Holliday junction formation (shown in the context of introducing an exogenous template in Figure 1.1). The removal of Holliday junctions and restoration of linear DNA molecules can result in either non-crossover and a perfect repair, or crossover, where the aligned chromatids swap portions of their DNA. Non-crossover is achieved by either dissolution by the mammalian BLM/TOP3a helicase-topoisomerase complex (Ira et al., 2003) or synthesis-dependent strand annealing (Nassif et al., 1994). Crossover events are caused by resolution of the junction by resolvases (endonucleases) like GEN1 (Ip et al., 2008), or those associated with the SLX4 complex (Munoz et al., 2009, Svendsen et al., 2009, Fekairi et al., 2009). For a detailed review of resolvases see West (2009). In the case of DSBs induced by genome editors, they are likely to be resolved by the gene conversion mechanism (see Svendsen and Harper (2010) for detailed description), as both ends of the DSB are present.

1.2.2. Meganucleases

Meganucleases are endonucleases which recognise and cleave DNA at specific sequences. Unlike restriction enzymes, which usually recognise a site of 3-8bp in length, meganuclease recognition sites can range in length from 12 to 44bp, but can tolerate some mismatches. This high specificity means that a meganuclease will only cut at a few loci within a genome; a favourable characteristic for a genome editor.

Meganucleases are represented in nature by homing endonucleases, an example being I-SceI, found in yeast mitochondria. These are mobile genetic elements which were first identified in yeast (Jacquier and Dujon, 1985) and are generally found within introns or inteins and are noted for their ability to home in on sequences which aren't already colonised by a homing endonuclease. There are four major families of homing endonuclease, named after amino acid motifs shared between the family members: LAGLIDADG or LHEs, which are the most studied homing endonuclease family, H-N-H, His-Cys box, and GIY-YIG (Stoddard, 2005). Homing endonucleases can induce DSBs in mammalian cells and, when used in conjunction with an exogenous template, can stimulate gene targeting via homologous recombination at

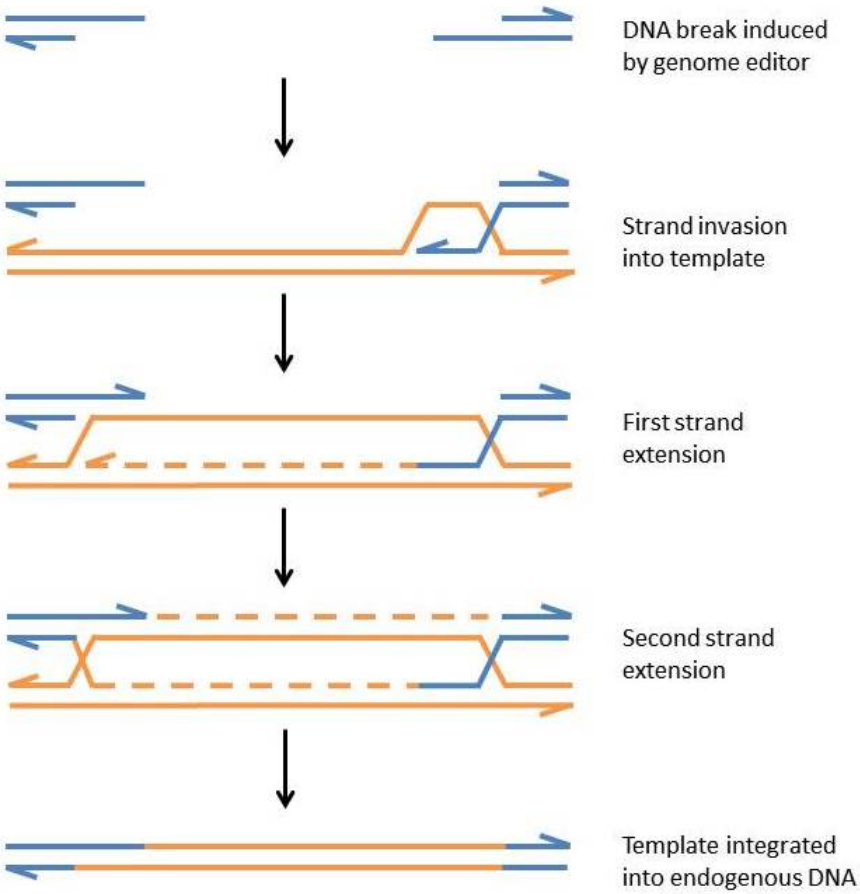


Figure 1.1.: Diagram showing repair of a DSB using an exogenous DNA template (orange) and Holliday junction formation.

a far greater efficiency than seen when the targeting vector alone is introduced (Rouet et al., 1994, Choulika et al., 1995). However, it is highly unlikely that a candidate gene for editing will contain a natural homing endonuclease recognition site. Therefore, the ability to change the recognition sequence of a homing endonuclease or to produce artificial meganucleases with relative ease is imperative to their use as genome editors.

The specificity of a meganuclease can be changed in a number of ways. If the meganuclease originates from a Group II intron, the specificity of the enzyme can be altered by substituting the intron RNA sequence which the endonuclease is associated with. This can be used to edit endogenous gene sequences by supplying the desired sequence as the intronic RNA (Guo et al., 1997, 2000). Changing the sequence specificity of a homing endonuclease originating from a Group I intron is more complex and requires the physical structure of the protein to be altered, such as the DNA-binding $\alpha\beta\beta\alpha\beta\beta\alpha$ folds in LHEs (Arnould et al., 2006, Smith et al., 2006, Silva et al., 2006). This can be done by domain swapping (Steuer et al., 2004) or domain fusion (Chevalier et al., 2002, Epinat et al., 2003), which exploits the natural DNA sequence specificity of other endonucleases, or by making point mutations within the enzyme (Seligman et al., 2002, Sussman et al., 2004). The majority of homing endonucleases which have been characterised so far have been isolated from Group I introns. The process of designing a custom meganuclease is made easier if the crystal structure of the natural endonuclease in complex with its DNA target is known (Flick et al., 1998, Jurica et al., 1998, Van Roey et al., 2001, Moure et al., 2003, Landthaler et al., 2006), but still necessitates the use of computational modelling if a specific DNA sequence is to be targeted (Ashworth et al., 2010), making the process of designing a custom meganuclease complex.

Despite this, meganucleases targeting endogenous DNA sequences have been produced (Smith et al., 2006, Grizot et al., 2009) and have also been used to correct deleterious alleles, such as those involved in SCID and Duchenne muscular dystrophy, with an efficiency of up to 6% (Arnould et al., 2007, Chapdelaine et al., 2010). They can also be used to create transgenic mammals capable of transmitting the transgene through the germline (Menoret et al., 2013, Wang et al., 2014c). It is also possible to engineer meganucleases with nickase activity (Niu et al., 2008, McConnell Smith et al., 2009), where only one DNA strand is cut. This reduces the chances of an NHEJ event occurring at the target site, while still providing favourable conditions for integration of exogenous DNA by HDR.

1.2.3. Zinc finger nucleases

Zinc finger domains were first described during characterisation of the transcription factor IIIA in *Xenopus* as repeats of about 30 amino acids in length which could bind zinc ions (Brown et al., 1985, Miller et al., 1985). The high frequency of cysteine and histidine residues within these repeats later led to these particular zinc finger domains being placed in the Cys_2-His_2 class of zinc fingers. Zinc finger domains which fall within this class consist of an α-helix and an anti-parallel β-sheet which, between them, bind a zinc ion (Pavletich and Pabo, 1991). DNA recognition is mediated by the α-helix, which binds in the major groove of the target DNA, and residues at the N-terminal end of the domain, which bind nucleotides on the same strand as those bound by the α-helix (Pavletich and Pabo, 1991). Each zinc finger domain binds to three consecutive nucleotides and different zinc finger domains recognise and bind to different DNA triplets (Huang et al., 1996).

This difference in binding specificities exhibited by different zinc finger domains is the basis on which a customisable zinc finger nuclease (ZFN) can be constructed from selected zinc fingers domains (typically three to four) attached to the FokI cleavage domain (Kim et al., 1996) (see Section 1.2.3.1). Due to FokI's need to function as a dimer, pairs of nucleases need to be produced which are designed to line up the two FokI domains at the same cleavage site (see Figure 1.2). Importantly, fusion of the FokI domain does not affect the DNA specificity of the zinc finger domains (Smith et al., 1999).

ZFNs can be used in gene targeting experiments to increase targeting efficiencies (Bibikova et al., 2001), resulting in efficiencies of 3-5% targeting in human cells (Porteus and Baltimore, 2003). They can also be used to stably correct endogenous genes in both immortalised cell lines and primary cells at a high efficiency, including a substantial population of cells with bi-allelic corrections (Urnov et al., 2005). The fact that ZFNs can be used in a variety of contexts, including in different species (Bibikova et al., 2003, Geurts et al., 2009, Watanabe et al., 2010, Hauschild et al., 2011, Young et al., 2011), as well as facilitating the formation of large insertions and deletions (Moehle et al., 2007, Lee et al., 2010), demonstrates the flexibility of the system. However, it is not currently possible to target any given DNA sequence as not all DNA triplets are recognised by a zinc finger domain. In addition to this, DNA binding by arrays of zinc finger domains is affected by context-dependent effects (i.e. positioning two particular

zinc finger domains next to each other can influence the strength of DNA binding) (Cathomen and Joung, 2008). These drawbacks mean there is scope for a more universal genome editor system to be developed.

1.2.3.1. FokI

FokI is a Type IIs restriction enzyme originally isolated from *Flavobacterium okeanokoites* (Li et al., 1992). Type IIs enzymes do not cleave DNA within the recognition site, but rather a precise distance away from the recognition site. Wildtype FokI consists of two subunits; a 41-kDa domain which recognises the target DNA sequence of 5'GGATG-3' (Hiroyuki and Susumu, 1981) and a 25-kDA domain which has endonuclease activity (Li et al., 1992). When two FokI molecules dimerise, the target DNA is cleaved nine bases downstream of the recognition site (equivalent to one turn in the DNA helix) (Hiroyuki and Susumu, 1981, Bitinaite et al., 1998) and FokI can cleave both methylated and unmethylated DNA (Li et al., 1992). In the absence of a target DNA sequence, the endonuclease domain is sequestered by the recognition domain, preventing off-target cleavage (Vanamee et al., 2001).

FokI has been used previously to create a chimeric restriction enzyme, by fusion of its endonuclease domain to the *Drosophila* Ubx homeobox domain. This changed the target DNA sequence of the enzyme to the sequence recognised by the Ubx domain (Kim and Chandrasegaran, 1994). To improve the cutting efficiency of enzymes using the FokI endonuclease domain, a linker is used to separate the endonuclease domain from the DNA recognition domain (Smith et al., 2000). In the case of ZFNs, the spacer is usually 18 amino acids long (Bibikova et al., 2001).

Chimeric enzymes containing the FokI endonuclease domain need to dimerise in order to function, as the FokI domain cannot function as a monomer (Bitinaite et al., 1998, Smith et al., 2000). In order for dimerisation to occur, the target DNA sequences must be present, as FokI only has a weak dimerisation interaction on its own (Vanamee et al., 2001). Enzyme dimers must also have a tail to tail conformation, in order for the two FokI domains to be brought together (Bitinaite et al., 1998).

Since its initial use as an enzymatic domain in genome editors, the endonuclease domain of FokI has been altered with a number of point mutations to increase the cutting efficiency and specificity of the editors. The mutations S418P and K441E (also known as the 'Sharkey' muta-

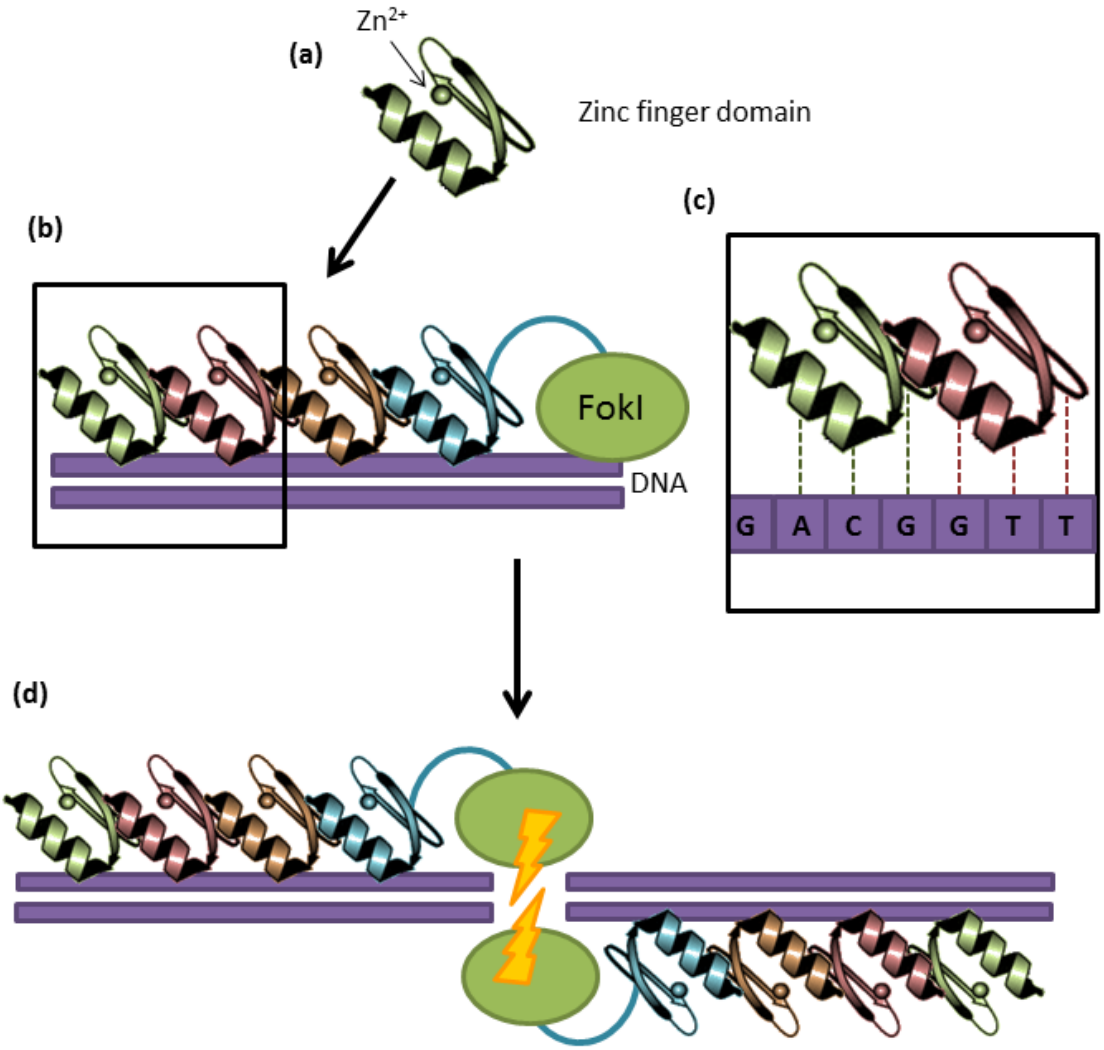


Figure 1.2.: Structure and function of zinc-finger nucleases. (a) Each zinc finger domain contain an α-helix and two β-strands with a zinc ion held between the α-helix and one of the β-strands. (b) Zinc finger nucleases consist of a series of zinc-finger domains connected to the FokI endonuclease by a spacer. Each zinc finger domain recognises a triplet of nucleotides, as shown in inset box (c), with the two domains recognising the triplets 'ACG' and 'GTT' respectively. (d) Zinc finger nucleases function as dimers, leading to the creation of a double strand break in the DNA. Image adapted from Gaj et al. (2013)

tions) confer an up to 15-fold increase in activity compared to wildtype FokI (Guo et al., 2010), while the production of obligate heterodimeric domains has helped to increase the specificity. These heterodimeric mutations are located within a hydrophobic pocket of FokI and are designed so that heterodimers contain oppositely charged amino acids within the pocket. the attraction between the amino acids in the heterodimers increase the likelihood of the heterodimeric complex forming, while also repelling any dimers with the same amino acids (Miller et al., 2007, Szczepek et al., 2007, Doyon et al., 2011). The most commonly used heterodimers have the point mutations Q486E, I499L and N496D (also known as ELD) and I538K, E490K, and H537R (also known as KKR) and their use can reduce cytotoxicity effects seen when genome editors are introduced into a cell or embryo (Young et al., 2011).

1.2.4. TALENs

Transcription activator-like effector nucleases (TALENs) are derived from the transcription activator-like effectors (TALEs) expressed by bacteria in the *Xanthomonas* genus, which are plant pathogens (Bonas et al., 1989). TALEs, such as AvrBs3, are secreted factors which enter the plant cell (Van den Ackerveken et al., 1996) and interact with the host genes via the TALE's DNA recognition domain. Activation of endogenous genes by TALEs can have the effect of either inducing changes in the cell which benefit the pathogen, such as hypertrophy of the cell to enhance the spread of the bacteria (Kay et al., 2007), or alternatively inducing a resistance response in the plant (Bonas et al., 1989).

The DNA recognition domain of TALEs is comprised of an array of TALE repeats, each of which is about 34 amino acids in length. The repeats have an almost identical amino acid sequence to each other, with the exception of the residues at positions 12 and 13 within each repeat. These two amino acids are known as the repeat variable diresidue and are found within a loop which connects the two α-helices in each TALE repeat (Boch and Bonas, 2010). The combination of RVDs within a TALE determine its DNA specificity (Bonas et al., 1989, Herbers et al., 1992), with each RVD recognising a single nucleotide in the target DNA (Boch et al., 2009, Moscou and Bogdanove, 2009).

TALENs are constructed from arrays of TALE repeats (usually around 18 RVDs in length) connected to a FokI cleavage domain via a spacer (Li et al., 2011, Christian et al., 2010). In comparison to ZFNs, TALENs' simple 1:1 nucleotide:RVD code and relative lack of context-

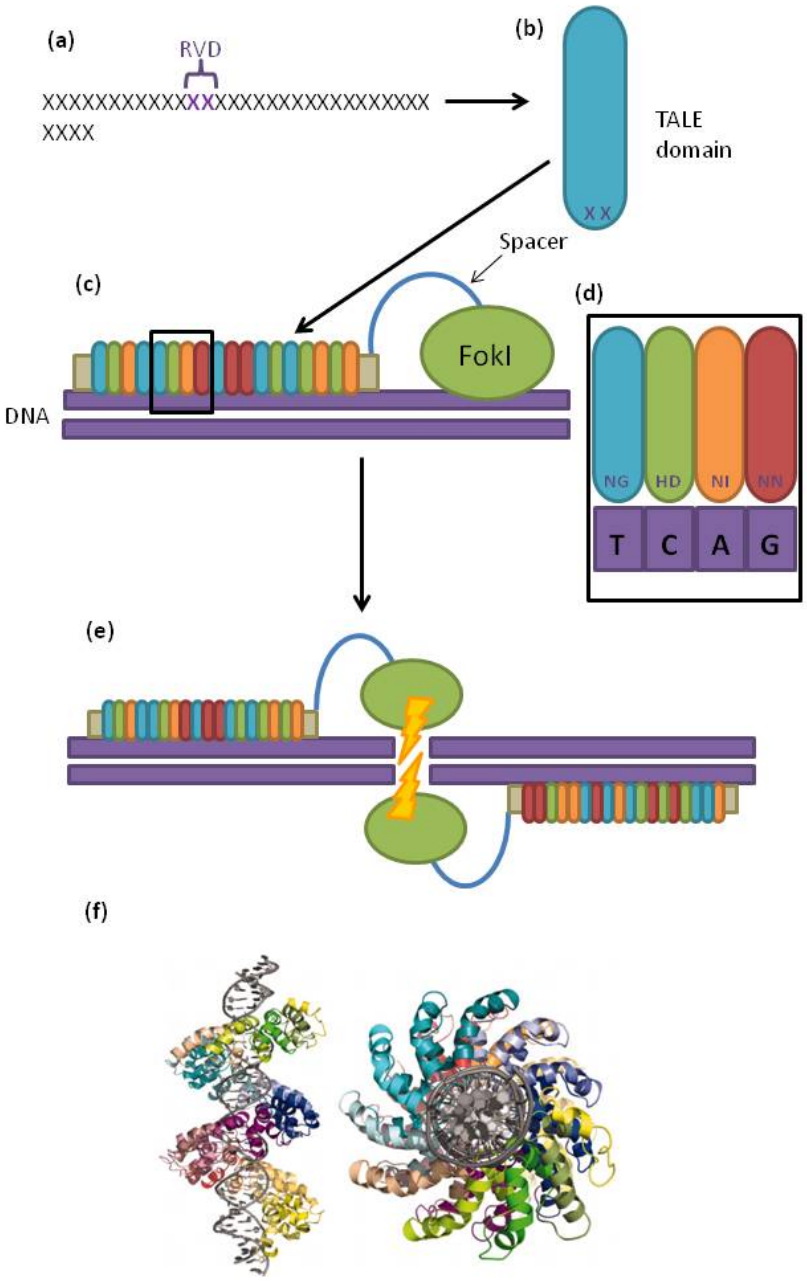


Figure 1.3.: Structure and function of TALENs. (a) Each TALE DNA-targeting domain contains 34 or 20 amino acids, making up a full-length or truncated repeat, respectively. The amino acids are arranged in tandem repeats, with polymorphic residues at position 12 and 13 forming the RVD (shown by the purple 'XX' in (a) and (b)). (b) The RVD is positioned within the TALE domain at the point of contact with the DNA strand. (c) Each TALEN contains 13-28 of these domains attached to a 3' FokI nuclease by a small spacer. The inset box (d) shows the 1:1 binding relationship between each subunit and a nucleotide on the targeted DNA strand. (e) FokI can generally only function as a dimer so two TALENs must bind to a stretch of DNA to induce a double-strand break between the two binding sites, however monomeric TALENs are becoming available (Kleinstiver et al., 2014). (f) Structure of the PthXo1 TALE bound to DNA reproduced from Mak et al. (2012).

dependent effects offer greater flexibility meaning that any DNA sequence within a genome can be targeted while sharing a similar mutagenic frequency to ZFNs (Bogdanove and Voytas, 2011). TALENs which use the FokI cleavage domain must been constructed and used as dimers in the same way as ZFNs (see Figure 1.3). This reduces the potential for off-target effects by providing greater specificity.

The ease of producing TALENs as compared to ZFNs allows them to have a wider range of applications, including the ability to create TALEN arrays which span an entire genome (Kim et al., 2013a) or TALENs which can edit miRNAs (Kim et al., 2013c, Takada et al., 2013). TALENs can also be inserted into an organism's genome as DNA under the control of an inducible promoter to create conditional knockouts (Cheng et al., 2013) and, like ZFNs, have been shown to function efficiently in a wide range of species (Carlson et al., 2012, Sung et al., 2013, Liu et al., 2013, Tan et al., 2013, Park et al., 2014).

1.2.5. CRISPRs

The CRISPR/Cas9 system represents a different form of genome editor compared to TALENs and ZFNs. Instead of a series of DNA-recognising protein domains physically attached to an endonuclease, CRISPR/Cas9 editors are made up of two separate molecules; a short guide RNA (sgRNA) molecule, which provides the sequence specificity and a Cas9 endonuclease. The Cas9 enzyme and sgRNA complex together to form the active editor.

CRISPRs, or clustered regularly interspaced short palindromic repeats, were first identified as a series of palindromic repetitive sequences within the genomes of bacteria (Ishino et al., 1987, Nakata et al., 1989). These repeats were alternated with 'spacer sequences' which showed no repetition or, indeed, any homology to the rest of the genome. Further analysis determined that the spacer sequences were extrachromosomal elements derived from invading pathogens, such as bacteriophages and plasmids (Bolotin et al., 2005, Barrangou et al., 2007, Garneau et al., 2010). This led to the identification of the CRISPR system as a heritable form of adaptive immune system which is present in both bacteria and archaea (Bhaya et al., 2011, Richter et al., 2012), where the spacer sequences are transcribed and processed into small CRISPR RNAs (crRNAs) by CRISPR associated (Cas) proteins, RNaseIII and trans-activating crRNAs (tracrRNAs) to invoke cleavage of invading pathogenic DNA (Brouns et al., 2008, Deltcheva et al., 2011). tracrRNAs also bind mature crRNAs via Watson-Crick base-pairing, a

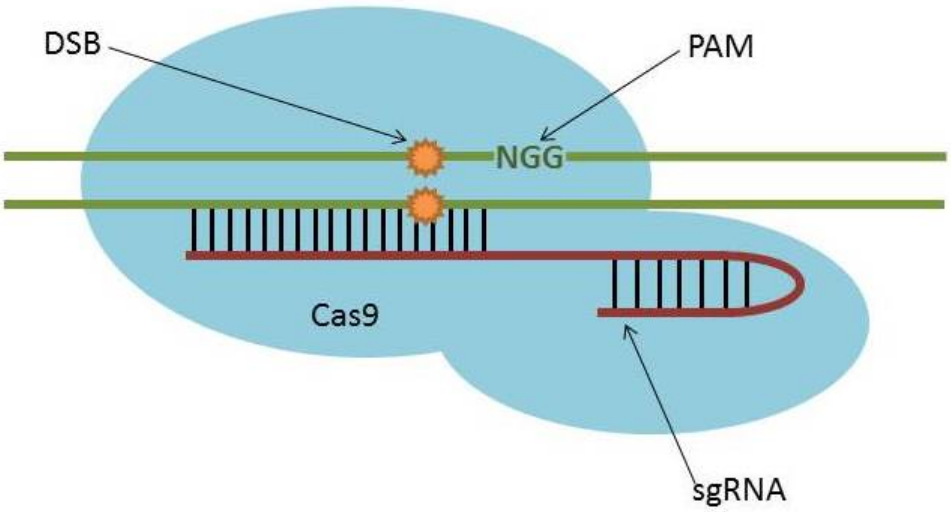


Figure 1.4.: Diagram of Cas9:sgRNA:target DNA complex. Target DNA containing the PAM sequence is shown in green. Orange stars indicate where the double strand break is induced in the DNA, upstream of the PAM.

structure required for successful cleavage by Cas9. The stem-loop structures in tracrRNAs also help to stabilise the final Cas:crRNA:tracrRNA:DNA complex (Nishimasu et al., 2014, Jinek et al., 2012).

To date, three types of CRISPR system have been identified (Makarova et al., 2011). The Type I and Type III systems make use of large multiprotein complexes known as Cascade complexes to process crRNAs and cleave foreign DNA, making them unwieldy for use as a genome editor at this time. However, the Type II system, such as the one found in *Streptococcus pyogenes*, relies on a single enzyme, Cas9, to carry out both crRNA maturation and DNA cleavage (Jinek et al., 2012) and is easily programmed to target a desired DNA sequence (Jinek et al., 2012). The system is simplified by the use of a chimeric short guide RNA (sgRNA), where the 3' end of a crRNA is fused to the 5' end of a tracrRNA. A target DNA sequence is specified by a 20bp sequence located in the crRNA portion of the sgRNA (Jinek et al., 2012) (see Figure 1.4).

Studies of the crystal structure of the Cas9 enzyme have shown the presence of two main lobes; one involved in target recognition and one with nuclease activity. The target recognition lobe contains a positively charged groove where the negatively charged sgRNA:target DNA complex is held, while the nuclease lobe cleaves target DNA through its RuvC and HNH domains. Each nuclease domain is capable of cutting a single DNA strand; the DNA strand

complementary to the sgRNA is cleaved by the HNH domain while the non-complementary strand is cut by the RuvC domain. The nuclease lobe also recognises the protospacer adjacent motif (PAM) on the non-target strand of DNA. Different CRISPR systems recognise different PAMs (Hou et al., 2013); in the case of Cas9 from the Type II system of *S. pyogenes*, the PAM sequence is NGG (Jinek et al., 2012). Recognition of a sequence complementary to the sgRNA must occur before the HNH domain can cleave the DNA (Jinek et al., 2012, Nishimasu et al., 2014). Once a suitable sequence has been identified, cleavage occurs 3bp upstream from the PAM (Jinek et al., 2012).

CRISPR/Cas9 has already been shown to be capable of the same range of editing events as ZFNs and TALENs, such as large deletions (Fujii et al., 2013, Han et al., 2014), gene corrections (Yin et al., 2014) and producing edited cells and animals (Jiang et al., 2013, Cho et al., 2013, Yang et al., 2013, Tan et al., 2013, Niu et al., 2014). However, one of the greatest advantages of the CRISPR/Cas9 system is that multiple sgRNAs for different targets can be introduced into a cell at the same time, a technique known as multiplex editing (Li et al., 2013, Cong et al., 2013). To this end, vectors have been developed to easily express multiple sgRNAs at the same time (Tsai et al., 2014, Guilinger et al., 2014). Multiplex editing allows researchers to use the system in genome-wide screens for either loss- or gain-of function studies (Koike-Yusa et al., 2014, Shalem et al., 2014, Wang et al., 2014b, Konermann et al., 2014, Gilbert et al., 2014), or to study multigenic diseases such as cancer (Platt et al., 2014, Xue et al., 2014). Multiple sgRNAs can also be used to induce large chromosomal deletions as, unlike ZFNs and TALENs which generate overhangs at the cut site (Kim et al., 2013b), the CRISPR/Cas9 system creates blunt ends and can leave a clean deletion between the two cleavage sites, without any NHEJ artefacts (Jinek et al., 2012). This can be used in adult animals to induce carcinogenic chromosomal rearrangements, creating more accurate cancer models (Blasco et al., 2014, Maddalo et al., 2014).

Unlike dimeric ZFNs and TALENs, the simplicity of the monomeric CRISPR/Cas9 system has raised concerns about its fidelity and, more specifically, the frequency of off-target effects it induces. Articles published shortly after the establishment of the CRISPR/Cas9 system as a genome editing tool have shown notable off-target effects. These studies have either tended to look at the effects of mismatches on cleavage of one particular target, such as GFP (Fu et al., 2013), or have targeted a gene with very high sequence similarity to others in the genome (e.g.

CCR5 and CCR2) (Cradick et al., 2013) and have shown that mismatches at the 5' end of the sgRNA are better tolerated by the Cas9 system than those at the 3' end near the PAM (Fu et al., 2013). Indeed, it may be that, in cases where a target is known to share an unusually high degree of sequence homology with another site in the genome, the increased fidelity of ZFNs and TALENs would be better suited to targeting these loci.

However, other research has produced a more global picture of off-target effects by using both analysis of predicted off-target sites and whole-genome sequencing (Cencic et al., 2014, Veres et al., 2014, Smith et al., 2014, Yang et al., 2014, Suzuki et al., 2014). These works have found a much lower level of off-target cleavage than was originally predicted by the initial experiments described above. By studying the effects of genome editors on endogenous loci, rather than using an exogenous construct, the context in which these editing events were studied could be considered to be more relevant to the intended uses of genome editors (e.g. for producing modified endogenous alleles in cells and embryos). In addition, experiments on the original CRISPR/Cas system in bacteria have shown the system's low tolerance for mismatches (Barrangou et al., 2007). Available assays for determining the rate of on- and off-target effects induced by genome editors have been recently reviewed by Hendel et al. (2015).

Several methods exist where the frequency of potential off-target effects can be reduced. Modifying the design of the sgRNA by truncating the 3' end of the sgRNA or adding two target-independent guanine residues to the 5' end of the sgRNA can significantly reduce the number of off-target editing events, but can also impact on on-target editing efficiencies (Pattanayak et al., 2013, Cho et al., 2014). Introducing truncations at the 5' end of the sgRNA, in effect reducing the length of the sgRNA:target DNA interface from 20bp to 17bp, increases the specificity of sgRNAs without compromising the on-target activity (Fu et al., 2014).

FokI-Cas9 fusions have been produced to form dimerisation-dependent nickases (Tsai et al., 2014, Guilinger et al., 2014). Unlike monomeric Cas9 nickases, which can still potentially have off-target effects, the dimeric nickases require two PAM sites in close proximity to each other in order to function. This added level of specificity has been shown in recent assays to reduce the number of off-target cleavage events seen in comparison to the monomeric CRISPR/Cas9 system (Wang et al., 2015), particularly when used in conjunction with 5'-truncated sgRNAs (Fu et al., 2014). Given that an NGG PAM sequence occurs on average once every 8bp in the human genome (Cong et al., 2013) means that the FokI/Cas9 nickases retain much of the

flexibility which makes the CRISPR/Cas9 system so attractive.

1.3. Applications of genetically modified livestock

Despite the relative lack of technologies available to create genetically modified large animals in comparison to rodents, genetically modified livestock species, such as pigs, cattle, chickens and sheep can be very valuable in a variety of settings.

The physiological similarities between livestock species and human make large animals a prime candidate for producing disease models which closely mimic disease progression in humans. So far, pigs have been primarily used for this and models of hereditary diseases such as cystic fibrosis and Huntington's (Rogers et al., 2008b,a, Yang et al., 2010) as well as noncommunicable diseases such as diabetes and atherosclerosis (Wolf et al., 2014, Wei et al., 2012) have already been produced. Genome editors, particularly the multiplex editing afforded by use of the CRISPR/Cas9 system, could allow large animal models of multigenic diseases such as cancer and heart disease to now be produced (see Khaled and Liu (2014) for a review of current mouse models of cancer). In a similar vein, large animals have the potential to be used as bioreactors to produce therapeutic molecules used in the treatment of diseases including diabetes, melanoma and α1-antitrypsin deficiency (Pohajdak et al., 2004, Lillico et al., 2007, Wright et al., 1991, Carver et al., 1992) as well as being modified to produce organs suitable for xenotransplantations (Ekser et al., 2012, Klymiuk et al., 2010).

Given their inherent value in agriculture, there is demand to produce modified livestock with greater productivity, an area where animal biotechnology has so far lagged behind that of plant biotechnology. The most prominent example of this is the AquAdvantage salmon (Devlin et al., 1994), which is currently undergoing FDA assessment to become the first genetically modified animal to be approved for human consumption. If approved, the AquAdvantage salmon could be the first of many animals modified to be more productive, including the phytase-expressing Enviropig, which digests phosphates with greater efficiency than wildtype pigs (Golovan et al., 2001) and transgenic cattle which produce high casein levels in their milk, improving the quality (Brophy et al., 2003). Another key way to improve productivity among livestock animals is to reduce losses from diseases such as bird flu, mastitis and prion diseases by inserting disease resistance mechanisms into the genome (Lyall et al., 2011, Wall et al., 2005, Richt et al.,

2007) and as the underlying mechanisms behind key livestock diseases are examined in greater details, more options for producing disease- resistant livestock will become available.

This project sought to evaluate the use of genome editors to efficiently produce genetically modified animals for a range of applications; increased productivity, increased disease resistance and chimeric animals for tissue production.

2. Materials and Methods

All chemicals were purchased from Sigma-Aldrich unless otherwise stated.

2.1. Molecular biology

2.1.1. Bacterial media and antibiotics

Spectinomycin was sourced from AppliChem and 10mg/ml stocks made in water. Ampicillin was purchased from Sigma-Aldrich and used to make a 50mg/ml stock in water. Bacteria were grown either in LB medium or on agar plates (made by the Central Services Unit in the Roslin Institute), containing the appropriate antibiotic where applicable. All spectinomycin-containing media and plates had a final concentration of 50μg/ml and all ampicillin-containing media and plates were made to a final concentration of 100μg/ml. Plates were made by pouring antibiotic-containing LB agar onto 90mm single vent petri dishes.

Where selection for β-galactosidase was required, plates were prepared by first warming at 37°C for at least 30 minutes, then spreading 16μl of 50mg/ml X-gal and 40μl of 0.1M IPTG onto each plate.

2.1.2. DNA quantification using the Nanodrop spectrophotometer

DNA was routinely quantified using the Nanodrop spectrophotometer (Nanodrop) and ND-1000 software. Samples are analysed by pipetting 2μl aliquots onto the sensor, lowering the arm to touch the sample and taking a spectral measurement. To begin quantification, deionised water was used to calibrate the Nanodrop, followed a 'blank' sample of 2μl of the solvent which the DNA was suspended in (e.g. deionised water or elution buffer). A 2μl aliquot of the DNA-containing sample was then loaded onto the machine and the concentration of DNA given in ng/μl as well as the 260/280 ratio and the absorption spectrum of the sample. If

multiple samples were to be quantified, the sensor was cleaned with deionised water and a lint-free cloth between samples.

2.1.3. Analysis of RNA using the Agilent Bioanalyser

All TALEN and CRISPR RNA was analysed on RNA 6000 Nano chips (Agilent) with the Agilent 2100 Bioanalyser prior to use. Before setting up the chip, all reagents in the RNA 6000 Nano kit were left at room temperature for 30 minutes to equilibrate. 1µl of dye concentrate was added to a 65µl aliquot of filtered gel matrix and vortexed until the dye was evenly mixed into the gel. The gel-dye mix was then centrifuged at 13,000g for 10 minutes. 9µl of the gel-dye mix was pipetted into the chip (see Figure 2.1) and the chip primed on the Priming Station. After priming, 9µl of gel-dye mix was added to each of the remaining wells marked 'G'.

5µl of RNA 6000 Marker was added to the ladder well and each sample well to be used in the assay. Any unused sample wells were filled with 6µl of marker. 2µl aliquots of all the samples to be analysed and a 1µl aliquot of RNA 6000 ladder were incubated at 70°C for 2 minutes to remove any secondary structure which could affect analysis. After denaturation, 1µl of the ladder and all samples were added to the appropriate wells and the chip vortexed on the IKA vortex mixer at 2400 rpm for 1 minute. All chips were vortexed and assayed on the Bioanalyser within 5 minutes of the samples being added to the chip.

The two cleaning chips were filled with 350µl of RNase-Zap and nuclease-free water. Before an assay was started, the electrodes were decontaminated by placing the RNase-Zap chip in the machine for 1 minute, followed by the nuclease-free water chip for 10 seconds. The RNA Nano chip was then placed into the machine and the data analysed using the 2100 Expert software. Assay results were checked to determine the integrity of the RNA and its concentration. These

Figure 2.1.: Image of an Agilent RNA 6000 Nano chip

results are presented in Appendix A4.

2.1.4. Transformation protocol

Unless otherwise stated, Subcloning Efficiency DH5α competent cells (Life Technologies) were used in transformations.

Competent cells were stored at -80°C and thawed on ice prior to use. For each transformation, 50µl of cells were aliquoted into a microcentrifuge tube and 2µl (1-10ng) of DNA was added. After adding the DNA, the pipette tip was moved through the thawed cells to distribute the DNA evenly. Reactions were incubated on ice for 30 minutes, then heat-shocked at 42°C for 30 seconds, then returned to ice for a further 2 minutes. 950µl of LB media was added to each transformation before incubation with shaking at 37°C for 1 hour. 50µl of each transformation was plated onto an agar plate containing the appropriate antibiotic and incubated at 37°C overnight.

2.1.5. Qiagen Miniprep protocol

Small-scale preparation of plasmids was carried out using the QIAprep Plasmid Spin kit (Qiagen) and the standard kit protocol, as described below;

3ml of overnight bacterial culture was spun at 3000g for 5 minutes and the supernatant removed. Cell pellets were resuspended in 250µl of Buffer P1 containing RNase A and LyseBlue reagent. 250µl of Buffer P2 was added to begin cell lysis and each tube was inverted 4-6 times until the solution was uniformly blue. After 5 minutes incubation, 350µl of Buffer N3 was added to neutralise the lysis reaction and the tubes inverted until no blue colour remained. The tubes were then spun at 13,000rpm in a microcentrifuge for 10 minutes to pellet the precipitated cell debris. To bind the plasmid DNA from the supernatant, the supernatant was spun through a filter cartridge at 13,000rpm for 1 minute and the flow-through discarded. The bound DNA on the filter was washed with 750µl Buffer PE and again spun at 13,000rpm for 1 minute then spun again to remove any remaining wash buffer. To elute the DNA, 50µl of Buffer EB was pipetted onto the filter and incubated at room temperature for 1 minute, before a final spin at 13,000rpm for 1 minute. DNA concentration was quantified using a Nanodrop.

2.1.6. Qiagen Endotoxin-free (EndoFree) Maxiprep protocol

Large-scale production of endotoxin-free plasmid was produced using the Qiagen EndoFree Plasmid Maxi kit. Competent cells were transformed with the plasmid of interest as in Section 2.1.4 and plated out. After overnight incubation at 37°C, one colony was picked and incubated in 3ml of LB media containing the appropriate antibiotic at 37°C in a shaking incubator for 6-8 hours. Overnight cultures were then set up with 100ml of antibiotic-containing LB media, to which the day culture was added and incubated overnight at 37°C.

Overnight cultures were spun at 6000g for 15 minutes at 4°C and the supernatant removed. To each cell pellet, 10ml of Buffer P1 containing RNase A and LyseBlue was added and the pellets resuspended by pipetting. The lysis reaction was started by the addition of 10ml of Buffer P2 and was incubated for 5 minutes before the reaction was stopped with 10ml of chilled Buffer P3. The solution was filtered into the QIAfilter cartridge through a layer of Miracloth (Merck Millipore) and was incubated at room temperature for 10 minutes. The lysate was then filtered through the QIAfilter cartridge into a 50ml Falcon tube. To remove endotoxins from the lysate, 2.5ml of Buffer ER was added to each sample and the tubes incubated on ice for 30 minutes. This step prevents endotoxins binding to the resin in the columns in the subsequent steps. During incubation, a QIAGEN-tip 500 column was equilibrated by applying 10ml of Buffer QBT and allowing it to flow through the column. After endotoxin removal, the lysate containing the plasmid of interest was poured into the column and allowed to flow through. The column was then washed twice with 30ml of Buffer QC before elution of the DNA in 15ml of Buffer QN into an endotoxin-free tube. 10.5ml of isopropanol was added to the tube to precipitate the DNA and the tube spun at 15,000g for 30 minutes at 4°C to pellet the DNA. The pellet was then washed in 5ml of endotoxin-free 70% ethanol and spun at 15,000g at 4°C for 10 minutes. The ethanol was removed and the pellet air dried in a microbiological safety cabinet for 5 minutes to allow any remaining ethanol to evaporate. The DNA pellet was then resuspended in 200µl of Buffer TE in the cabinet and incubated overnight at 4°C to allow the DNA to fully resuspend. Concentration of DNA was then quantified using a Nanodrop.

2.1.7. PCR

All PCRs were carried out using DreamTaq 2x mix (Thermo Scientific). A 10µl PCR is made up using 5µl of DreamTaq mix, 2.6µl of water and 0.2µl of both the forward and reverse primers. 2µl of template DNA was added to the 8µl of master mix before the PCR was run on a TProfessional thermocycler (Biometra). Unless otherwise stated, programs followed a standard design, as shown below. Annealing temperatures of PCR primers are given in Appendix A.1.

95°C	2 minutes	x1
95°C	30 seconds	x30
Annealing temperature	30 seconds	
72°C	1 minute	
72°C	10 minutes	x1

If required, PCR reactions were cleaned up using either the illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) or the ChargeSwitch kit (Life Technologies). Brief descriptions of the protocol for both kits are given below.

illustra GFX kit 500µl of Capture buffer type 3 was added to a 50µl PCR reaction and mixed thoroughly by pipetting. The colour of the mixture was then checked to ensure that it had not turned pink or red, which would indicate that the pH of the mixture would be too high for efficient isolation of the PCR product.

Mixtures were loaded onto a GFX MicroSpin column in a collection tube and centrifuged at 16000g for 30 seconds. The flowthrough was discarded and the column washed with 500µl of Wash buffer type 1 and centrifuged at 16000g for 30 seconds. Again, the flowthrough was discarded and the purified PCR product eluted in 20µl of Elution buffer type 6 by incubation for one minute and centrifugation at 16000g for one minute.

ChargeSwitch kit The ChargeSwitch magnetic beads were vortexed to fully resuspend the beads in the buffer. 50µl PCR reactions were mixed with 50µl of Purification Buffer by pipetting and 10µl of magnetic beads added and mixed by gentle pipetting until the beads were uniformly spread through the solution.

Solutions were incubated for one minute then placed into the MagnaRack station for one

minute until the beads formed a tight pellet against the wall of the tube. The supernatant was removed to leave the beads in the tube and the beads washed by resuspension in 150μl of Wash Buffer and placed in the MagnaRack station for one minute until a pellet had formed. Used wash buffer was removed without disturbing the pellet of beads and the beads washed once more.

Purified PCR product was eluted by mixing the beads in 25μl of Elution Buffer and incubating for one minute. The mixtures were then placed in the MagnaRack station for one minute or until the majority of the beads had been removed from the buffer. Buffer containing PCR products was carefully removed from the tube and stored at -20°C.

2.1.8. Gel electrophoresis

Gel electrophoresis was used to assess the products of PCR or restriction digests. Tris-acetate-EDTA (TAE) buffer was used for all electrophoresis. Agarose gels were made using UltraPure agarose (Life Technologies). Unless stated, gels were run using the GeneRuler 1kb DNA ladder (Thermo Scientific) as a size marker in the left hand lane.

2.1.9. DNA sequencing

Routine DNA sequencing of PCR products was carried out by Edinburgh Genomics, who carried out both the Big Dye sequencing reactions (Life Technologies) and the subsequent clean-up and capillary analysis using an ABI3730 DNA Analyser (Applied Biosystems). 3-10ng of cleaned-up PCR product (see Section 2.1.7) was sent to Edinburgh Genomics in 5μl of deionised water with 1μl of sequencing primer at a concentration of 3.2ng/μl. Sequencing data from Edinburgh Genomics was returned as .ftv files and initially analysed using FinchTV (Geospiza Inc), with subsequent work carried out using the Lasergene software package (DNASTAR) as necessary.

On occasion, it was necessary to clone PCR products prior to sequencing (e.g to confirm the nature of an editing event). PCR products were cloned using the CloneJET PCR Cloning Kit (Thermo Scientific) using the sticky-end cloning protocol. Firstly, a blunting reaction was set up containing 10μl of 2x Reaction Buffer, 1μl of PCR product and 1μl of the DNA Blunting Enzyme. The total reaction volume was made up to 18μl with nuclease-free water. The reaction mixture was vortexed and briefly centrifuged before incubation at 70°C for 5 minutes. After

incubation, the reaction was chilled on ice. To ligate the blunted PCR products into a plasmid vector, 1µl of the pJET1.2/blunt Cloning Vector (at 50ng/µl) and 1µl of T4 DNA Ligase were added to the completed blunting reaction. Again, the reaction was vortexed and centrifuged before incubation at room temperature for 5 minutes.

Ligation products were then directly used for bacterial transformation (see Section 2.1.4) or stored at -20°C until needed. Transformed colonies were picked for overnight cultures and the plasmids purified using the Miniprep protocol detailed in Section 2.1.5. 200-500ng of plasmid with 1µl of sequencing primer (at 3.2ng/µl) in a total volume of 6µl was sent to Edinburgh Genomics for sequencing as described above.

2.2. TALEN construction

2.2.1. TALEN design

TALENs were designed using the ZiFit designer software (`http://zifit.partners.org/ZiFiT/`). Briefly, a segment of DNA sequence of 300-400bp in length which surrounded the intended TALEN target was entered into the software. Where appropriate, the target site was indicated by the use of square brackets surrounding the nucleotide(s) of interest. Following analysis of the DNA sequence by the software, three TALEN pairs were selected based on proximity to the target site and were assembled.

2.2.2. Assembly of TALEN plasmids

TALEN plasmids were constructed using the GoldyTALEN scaffold (Bedell et al., 2012) and the Golden Gate cloning protocol (Cermak et al., 2011), which is described below.

The first round of cloning constructed arrays of up 10 consecutive RVDs in pFUS vectors. Different pFUS vectors are required to build TALENs of different lengths. For TALENs of up to 21 RVDs, the plasmid pFUS_A is used for RVDs 1 - 10, while the pFUS_B vectors are used for RVDs 11 - n-1 (Table 2.1). pFUS_B plasmids are numbered 1-10, indicating the number of RVDs which can be cloned into them (e.g. pFUS_B7 can hold 7 RVDs). This does not include the last RVD in the sequence, which is added in the second round of cloning.

For TALENs more than 21 RVDs in length, the plasmid pFUSA30A is used for RVDs 1-10 and the plasmid pFUSA30B for RVDs 11-20. The remaining RVDs (21 onwards) are inserted

Vector name	Purpose
pFUS_A	Vector for RVDs 1-10 of a TALEN up to 21 RVDs long
pFUSA30A	Vector for RVDs 1-10 of a TALEN over 21 RVDs long
pFUSA30B	Vector for RVDs 11-20 of a TALEN over 21 RVDs long
pFUS_B(1-10)	Vector for RVDs 11-(n-1) of a TALEN up to 21 RVDs long
pRCIScript-GoldyTALEN	Vector for full-length TALENs
pCAG-T7-TALEN(Sangamo)-FokI-ELD	Vector for full-length TALENs
pCAG-T7-TALEN(Sangamo)-FokI-KKR	Vector for full-length TALENs

Table 2.1.: Table showing all plasmid vectors used in TALEN construction

into the appropriate pFUS_B vector as above. Plasmid maps are shown in Appendix A.3. All destination vectors for full-length TALENs carry the Sharkey mutations described in Section 1.2.3, where the ELD and KKR heterodimeric mutations are also explained.

The following protocol is written assuming the construction of a TALEN which is 18 RVDs long, the length of the majority of the TALENs produced for this project.

To begin TALEN assembly, 150ng of each RVD (at a concentration of 150ng/µl) required was added to a PCR tube along with 150ng of the appropriate destination vector (i.e. pFUS_A or pFUS_B). To each reaction, 10U of BsaI restriction enzyme, 40U of T4 DNA ligase and 2µl of 10x T4 DNA ligase buffer (all NEB) were added and the total reaction volume was made up to 20µl with water. The example below shows the reagents used in a reaction to create a stretch of 10 RVDs in a pFUS_A vector.

Reagent	Volume
RVD1	1μl
RVD2	1μl
RVD3	1μl
RVD4	1μl
RVD5	1μl
RVD6	1μl
RVD7	1μl
RVD8	1μl
RVD9	1μl
RVD10	1μl
pFUS_A	1μl
10x DNA ligase buffer	2μl
BsaI	1μl
T4 DNA ligase	1μl
Water	5μl

The reactions were placed in a PCR machine and the following cycle was run. This cycle allows simultaneous digestion of individual RVD plasmids and ligation of digested plasmids into the completed subset.

37°C	5 minutes	10x
16°C	10 minutes	
50°C	5 minutes	1x
80°C	5 minutes	1x

After digestion/ligation, 1μl of 10mM ATP and 1μl of PlasmidSafe nuclease (Cambio) was added to each reaction. The reactions were incubated at 37°C to remove any unligated linear DNA molecules. The reactions were then transformed into competent cells using the protocol described in section 2.1.4, increasing the amount of DNA added to each transformation to 5μl. Agar plates containing spectinomycin were prepared for β-galactosidase selection as described in Section 2.1.1. Before plating, cells were spun at 13,000 rpm in a microcentrifuge for 3 minutes and the majority of the supernatant poured off. The cell pellet was resuspended in the

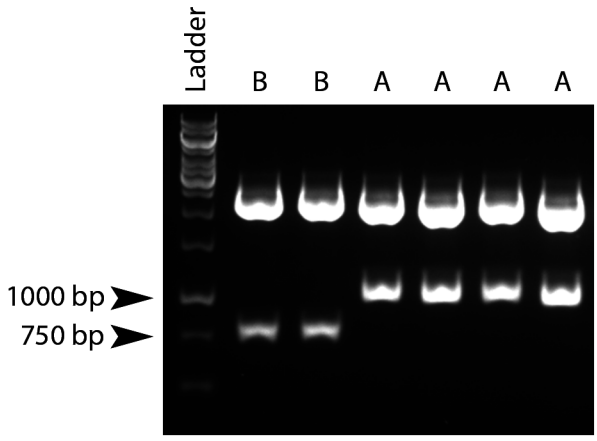


Figure 2.2.: Gel image showing correctly digested pFUS_A and pFUS_B7 vectors with RVD inserts. Lanes labelled 'B' contain pFUS_B7 constructs, while those labelled 'A' contain pFUS_A inserts

remaining supernatant (~200µl) and 50µl plated out followed by incubation at 37°C overnight. Up to four colonies from each plate were picked and overnight cultures were set up in 3ml of LB media containing 50ng/ml spectinomycin. Cultures were incubated at 37°C with shaking overnight. Plasmids of the RVD subsets were purified from the cultures as described in Section 2.1.5.

To test that the purified plasmids had inserts of the correct length, DNA was digested with XbaI (Promega) and AflII (NEB). This combination of restriction enzymes excises the RVD insert from the destination vector. Each digestion contained 5µl of DNA (at approximately 150-300ng/µl), 10U of XbaI, 20U of AflII, 1µl of 10x Buffer 4 (NEB) and 2µl water. Reactions were incubated at 37°C for one hour, then the products assessed by gel electrophoresis. Reactions were run out on a 1.5% agarose gel at 10V/cm for 60 minutes. Correctly assembled pFUS_A subsets produced a band of 1048bp, while pFUS_B7 subsets showed a band of 740bp (Figure 2.2).

To assemble the full-length TALEN, correctly assembled pFUS plasmids were diluted to 150ng/µl. 1µl of each subset was pipetted into a PCR tube and 1µl of the final RVD (at 150ng/µl) was added. To this, 75ng of the final destination vector (either pRCI-Script, pCAG-T7-FokI-ELD or pCAG-T7-FokI-KKR), 10U of Esp3I (Thermo Scientific), 40U of T4 DNA ligase and 2µl of 10x DNA ligase buffer were added and the final reaction volume increased

to 20µl with water. The reactions were placed in the PCR machine and the following digestion/ligation cycle run.

37°C	5 minutes	5x
16°C	10 minutes	
37°C	15 minutes	1x
80°C	5 minutes	1x

The reactions were transformed into competent cells and plated onto ampicillin plates with X-gal and IPTG. Plates were incubated at 37°C overnight and three colonies from each plate were picked to set up overnight cultures. Cultures were incubated in LB with ampicillin overnight with shaking at 37°C and the plasmid purified from cells using the Qiagen miniprep kit.

Purified full-length TALENs were subjected to two separate digests; a digest with StuI and AatII (NEB) to check that the TALEN insert was of the correct length, giving bands of 360bp, 1750bp, 2000bp and 3575bp for an 18RVD TALEN (Figure 2.3) and a digest with BspEI (NEB), which cuts in HD repeats only, giving bands in 100bp increments, depending on the RVD sequence (Figure 2.4). This BspEI digest gave an initial indication whether the RVD sequence was correct, although all TALENs were subsequently checked by sequencing (Eurofins MWG) to verify that the RVD sequence was correct. In the case of TALENs greater than 18 RVDs in length, the pFUS subsets, rather than the full-length TALEN, were sent for sequencing in order to get adequate overlap between the forward and reverse sequences. Correctly assembled TALEN plasmids were then expanded using the EndoFree Maxiprep protocol detailed in Section 2.1.6.

2.2.3. TALEN RNA synthesis

Restriction digests to linearise TALEN plasmids prior to RNA synthesis were set up using 5µg of TALEN plasmid, 0.5µl of 10U/µl of SacI and 10µl of 10x buffer A (Roche). Nuclease-free water (Ambion) was added to each digestion reaction to a total volume of 100µl. Reactions were then incubated at 37°C for 2 hours before being terminated by the addition of 5µl of 0.5M EDTA at pH 8.0, 10µl of 3M sodium acetate at pH 5.2 and 200µl of 100% ethanol. Linearised DNA was precipitated by incubating at -20°C for 15 minutes then centrifuging at 16,000g for

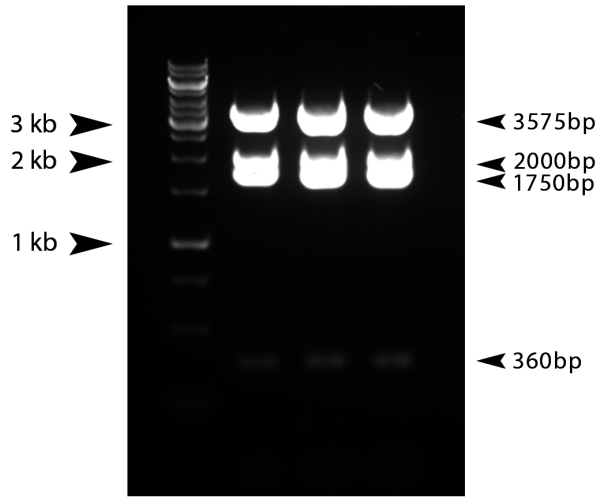
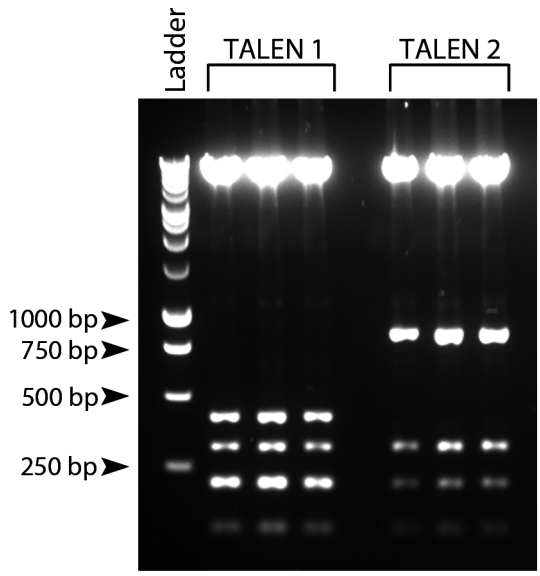


Figure 2.3.: Gel image showing correctly assembled TALENs after StuI/AatII digest. Sizes of bands in the GeneRuler 1kb DNA ladder are given on the left side of the image, while the expected band sizes resulting from StuI/AatII digest are shown on the right

15 minutes at 4°C. The supernatant was removed and the DNA pellet resuspended in 10µl of nuclease-free water. Once the pellet was fully resuspended, the DNA concentration was checked on the Nanodrop before use in the transcription reaction.

mMessage mMachine kits (Ambion) were used to set up *in vitro* transcription reactions. For TALENs assembled into the RCIScript vector, the T3 kit was used, while the T7 kit was used for all other TALEN vectors. 1µg of linearised DNA was added to 10µl 2x NTP/CAP mixture and 2µl of 10x Reaction Buffer. 2µl of Enzyme Mix was then added to the reaction and the total reaction volume made up to 20µl with nuclease-free water. Reactions were incubated at 37°C for 2 hours before treatment with 1µl of TURBO DNase for 15 minutes at 37°C. Reactions were stopped by the addition of 15µl of Ammonium Acetate Stop Solution (5M ammonium acetate and 100mM EDTA) and 115µl nuclease-free water.

TALEN RNA was purified by phenol/chloroform extraction. To each reaction, 150µl of phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma-Aldrich) was added, then 150µl of chloroform. The aqueous phase was pipetted into a new nuclease-free microcentrifuge tube and the RNA precipitated with 150µl isopropanol. Precipitation was carried out at -20°C for 15 minutes and the RNA pelleted by centrifugation at 16,000g for 15 minutes at 4°C. RNA was resuspended in 20µl of nuclease-free water and analysed on the Agilent Bioanalyser as described in Section 2.1.3.



TALEN 1 RVD sequence																	
X	HD	X	X	HD	X	HD	X	HD	HD	X	X	X	HD	X	HD	HD	HD
	300 bp					200 bp				400 bp					100 bp		
				200 bp				100 bp				200 bp					

TALEN 2 RVD sequence																	
X	HD	HD	X	HD	X	X	X	X	X	X	X	HD	X	X	HD	X	HD
	100 bp			800 bp													
		200 bp										300 bp					

Figure 2.4.: Gel image showing BspEI digest of triplicates of two different TALENs. Tables showing the position of HD repeats within each TALEN sequence (other RVDs are shown by 'X'). The predicted band sizes are shown below each sequence. Note that BspEI does not cut in HD1 RVDs or at the final HD repeat.

2.3. CRISPR construction

2.3.1. sgRNA design

sgRNAs for use in the CRISPR/Cas system were designed using Zifit. As with the design of TALENs, a 300-400bp DNA segment surrounding the intended target site was entered into the design tool and a list of candidate sgRNA target sequences was returned. Candidate target sequences were first selected based on their proximity to the intended target site and then screened for potential off-target effects using the BLAST nucleotide search (`http://blast.ncbi.nlm.nih.gov`). Prior to searching for potential off-target hits, each sgRNA target sequence had 'NGG' added to the 3' end, replicating the PAM sequence necessary for cleavage. Four sgRNAs were designed for each locus targeted.

2.3.2. RNA from gBlocks

To produce sgRNAs for embryo injections, total sgRNA sequences, including the specific gene targeting sequence were ordered as gBlocks (Integrated DNA Technologies), as the pSpCas9(BB) vector was not immediately available. gBlocks were designed containing an individual sgRNA expressed from a U6 promotor. Primers were also ordered to amplify the sgRNA sequence from the gBlock and add a T7 promoter at the 5' end. PCR was carried out using these primers and the gBlock as a template. Purified PCR products from these reactions was then used as a template for sgRNA production using the MEGAshortscript T7 kit (Life Technologies). Transcription reactions were assembled at room temperature as follows;

Reagent	Volume (µl)
T7 10x Reaction Buffer	2
T7 ATP Solution (75 mM)	2
T7 CTP Solution (75 mM)	2
T7 GTP Solution (75 mM)	2
T7 UTP Solution (75 mM)	2
Template DNA (maximum of 150nM)	<8
T7 enzyme mix	2
Nuclease-free water	Up to 20µl total reaction volume

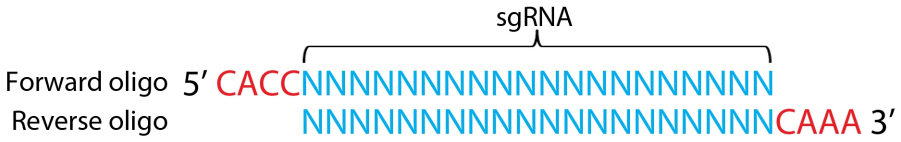


Figure 2.5.: Template for designing synthesised oligos for cloning of sgRNAs into the pSpCas9(BB) vector. Blue nucleotides show where the specific sgRNA sequence should be placed, while red nucleotides show the sticky ends added to facilitate ligation of the oligos into the BbsI restriction site in the plasmid.

Reactions were mixed gently and incubated at 37°C for 3-4 hours. After incubation, 1µl of TURBO DNase was added to each reaction and incubated at 37°C for 15 minutes to remove the DNA template. sgRNA was then recovered by phenol:chloroform extraction and ethanol precipitation as described in 2.2.3, resuspended in nuclease-free water and analysed on the Agilent Bioanalyser as described in Section 2.1.3.

2.3.3. Linker plasmid ligations

sgRNAs were cloned into the pSpCas9(BB) vector (see Appendix A.3) according to the protocol detailed in Ran et al. (2013). The pSpCas9(BB) plasmid contains a U6 promotor, which drives expression of the cloned sgRNA, the Cas9 protein and GFP. Briefly, forward and reverse linker oligos containing the 20bp targeting sequence for the sgRNAs and short overhangs matching the BbsI restriction site in pSpCas9(BB) were designed as shown in Figure 2.5 and sourced from Invitrogen.

Synthesised oligos were resuspended in deionised water to a final concentration of 100µM and the following reaction mix prepared for phosphorylation and annealing of the oligos. Reactions were carried out in a PCR machine at 37°C for 30 minutes followed by 95°C for 5 minutes. Reactions were then cooled to 25°C at a rate of 5°C a minute.

Reagent	Volume (µl)
Forward oligo	1
Reverse oligo	1
10x T4 ligation buffer	1
T4 PNK	1
Deionised water	6
Total	**10**

Phosphorylated and annealed oligos were diluted 1:200 in deionised water and cloned into the pSpCas9(BB) vector in the following ligation reaction;

Reagent	Volume (µl)
pSpCas9(BB) (1µg/1µl)	0.1
Diluted oligos	2
10x Tango buffer	2
10mM DTT	1
10mM ATP	1
FastDigest BbsI	1
T7 ligase	0.5
Deionised water	12.4
Total	**20**

Digestion/ligation reactions were incubated at 37°C for 5 minutes, followed by 21°C for 5 minutes. These two steps were repeated for a total of six cycles before completed reactions were transformed into competent cells as described in Section 2.1.4, plated onto ampicillin-containing LB plates and incubated at 37°C overnight. Individual colonies were picked and added to 3ml of ampicillin-containing LB medium and incubated at 37°C overnight with shaking. Plasmid DNA was purified from overnight cultures using the Qiagen Miniprep kit as detailed in Section 2.1.5. Purified DNA was stored at -20°C.

2.4. Transfection of cells with designer nucleases

Bovine embryonic fibroblasts (BEFs), ovine embryonic fibroblasts (OEFs) and an immortalised porcine kidney cell line (PK15) were used for validation of designer nucleases. All cells were cultured at 37°C and 5% CO_2 in high glucose DMEM containing GlutaMax and sodium pyruvate (Life Technologies, 31966-021) with 10% FCS (Sigma-Aldrich) and penicillin/streptomycin (Life Technologies). Cells were passaged by removing the medium with an aspirator and removing any remaining serum by washing in PBS. Cells were then bathed in TrypLE Express (Life Technologies) and incubated at 37°C until cells had dissociated from the plastic. Medium was added to each flask to quench the trypsin and the dissociated cells collected into a 15ml Falcon tube. Cells were centrifuged at 500g for 3 minutes and the resulting pellet split between new flasks or plates as appropriate.

Cells were transfected using the Neon electroporation system (Life Technologies) with 100µl reaction tips. To assemble the reactions, cells were trypsinised, pelleted and resuspended in PBS prior to further centrifugation. Cells were resuspended in PBS and a 10µl aliquot of the cells was taken to determine cell concentration using a haemocytometer. The total number of cells needed for all reactions were then aliquoted into a new tube and pelleted, meaning that the cells had now been washed twice in PBS.

To avoid the formation of bubbles within the pipette tip, which would disrupt the electroporation, a 150µl mixture of cells and nuclease was prepared for each transfection and 100µl of this taken for electroporation. For each reaction, 900,000 cells were resuspended in 150µl of Buffer R and 1.5µg of designer nuclease added to each aliquot of cells. The electroporation station was assembled according to the system protocol and 3ml of Buffer E2 was added to the Neon Tube in the Pipette Station. 100µl of each reaction, containing 600,000 cells and 1µg of designer nuclease, was taken up in the Neon pipette tip and electroporated using the following conditions;

Cell type	Voltage	Duration	Number of pulses
BEF and OEF	1800V	20ms	1
PK15	1400V	20ms	2

Each Neon tip was used for up to three electroporations before disposal. Figure 2.6 shows that three transfections of a GFP plasmid into OEFs carried out using the same Neon tip show

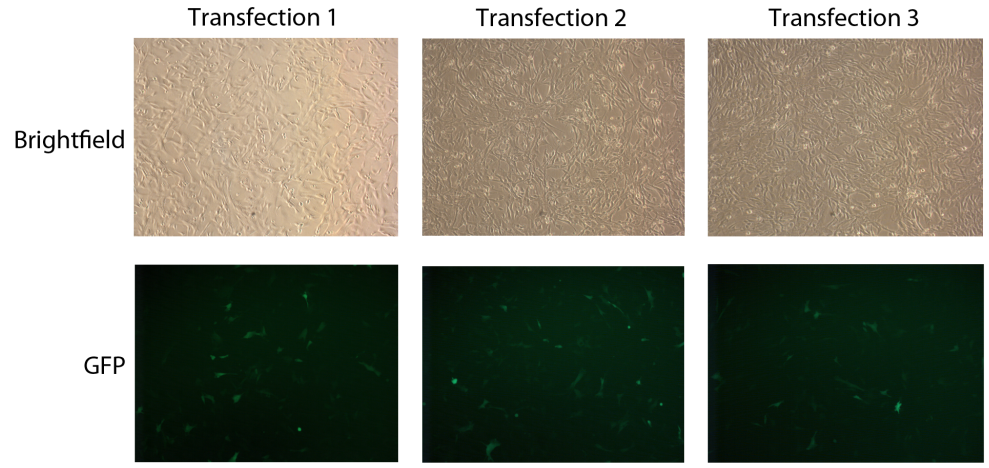


Figure 2.6.: Brightfield and GFP images of OEFs transfected with a GFP plasmid using the same Neon tip

similar transfection efficiencies. Following information in Tan et al. (2013), different incubation conditions were used depending on the form of nuclease transfected. Transfected cells were incubated for three days after electroporation according to the type of nucleic acid transfected, shown in the table below.

Nucleic acid	Incubation conditions
RNA only	30°C for three days
DNA only	37°C overnight, then 30°C for two days
RNA designer nuclease + GFP plasmid	30°C for two days, then 37°C overnight
DNA designer nuclease + GFP plasmid	37°C for three days

After three days, cells were harvested and pelleted by centrifugation at 200g for 5 minutes and removing the remaining medium. Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen). Cell pellets were resuspended in 200µl of PBS before the addition of 20µl of Proteinase K. 200µl of Buffer AL and 200µl of 100% ethanol were added to each sample and the resulting solution pipetted into a spin column. Columns were centrifuged at 16,000g for 1 minute to bind DNA to the column. The columns were then washed with 500µl of Buffers AW1 and AW2 and the DNA was eluted in 200µl of Buffer AE. DNA concentration was measured on the Nanodrop and designer nuclease activity assayed using either CelI or T7 endonuclease.

2.5. Analysis of genome editor activity using the Cel1 or T7 endonuclease assay

For both methods of analysis, the stretch of DNA targeted by the nuclease of interest was amplified by PCR. After PCR, the DNA underwent a denature/anneal step to allow the formation of heteroduplex DNA molecules. This was carried out in a PCR machine by heating the samples to 95°C for 10 minutes, then decreasing the temperature to 85°C at a rate of -2°C/sec. Reactions were held at 85°C for 1 minute and the temperature decreased by a further 10°C at a rate of -0.3°C/sec before pausing for 1 minute. This rate of decrease, with 1 minute holds at every 10°C was continued until the reactions reached 25°C.

For analysis using CelI, annealed PCR samples were run on a 1.5% agarose gel with the GeneRuler ladder to estimate DNA yield. 200-400ng of PCR product in a volume of 8-40µl was used in CelI digests with the SURVEYOR mutation detection kit (Transgenomic). The CelI enzyme in the kit cleaves DNA at points of mismatch in heteroduplexes. To each digest, 1/10th of the volume of MgCl was added, along with 1µl of Enhancer and 1µl of Nuclease. Digests were incubated at 42°C for an hour then run out on a 2% gel. Because the position of the nuclease cleavage site relative to the PCR product is known in each instance, the result of a successful CelI digestion can be predicted. In each case the presence of a full-length WT band and cleavage products were assessed.

For analysis using T7 endonuclease, PCR reactions were cleaned using the ChargeSwitch kit detailed in Section 2.1.9 before being run out on a 1.5% gel to quantify DNA in each reaction. A maximum of 200ng of PCR product was resuspended in 1x Buffer 3 (NEB) to a maximum volume of 19µl. 1µl of T7 endonuclease (NEB) was added to each 19µl solution and the reactions incubated at 37°C for 15 minutes. After incubation, reactions were immediately run on a 2% gel and the presence of cleavage products assayed.

For both the CelI and T7 assays, a positive control from the CelI assay kit was used. This positive control consists of two pre-made PCR reaction mixes containing a known template. The two templates differ from each other at one SNP, where one template has C and the other has a G. Both templates were amplified in individual 50µl PCR reactions. Upon completion of the PCR, 25µl was taken from each reaction and mixed together in a new tube before undergoing the denature-anneal process and the remaining steps of the assay with the other samples.

Because of the differing G/C SNP between the two templates, the denature-anneal process creates heteroduplexes which are recognised and cleaved by the CelI enzyme or T7 endonuclease, providing a positive control for the assay.

2.6. Embryo injection of designer nucleases

Preparation and transfer of embryos, as well as all animal husbandry was carried out by the staff at the Large Animal Unit. Embryo microinjections were carried out by Mr David Davies.

2.6.1. Preparation and microinjection of porcine embryos

The following media were prepared in advance of harvest and microinjection of porcine embryos.

NCSU-23 HEPES The following components were dissolved in 80ml of sterile water to produce a 100ml stock solution;

Reagent	Quantity
Sodium chloride	0.6355g
Sodium bicarbonate	0.0356g
Monopotassium phosphate	0.0162g
Magnesium sulphate heptahydrate	0.0293g
D-glucose	0.1g
L-glutamine (200mM)	0.5ml
HEPES (H-6147)	0.2383g
Taurine	0.0876g
Hypotaurine	0.0546g
Insulin (10mg/ml)	0.5ml
Penicillin G	0.0065g
Streptomycin sulphate	0.005g

0.025g of calcium chloride was then dissolved separately in 5ml of sterile water before being added to the stock solution. The resulting solution was then checked to ensure that the pH was

7.4 and the osmolarity fell between 260-280 mOsmol before being stored at 4ºC for up to two weeks.

NCSU-23 Base Medium The following reagents were dissolved in 150ml of sterile water;

Reagent	Quantity
Sodium chloride	1.2710g
Sodium bicarbonate	0.4212g
Potassium chloride	0.0713g
Monopotassium phosphate	0.0324g
Magnesium sulphate heptahydrate	0.0586g
D-glucose	0.2002g
L-glutamine (200mM)	1ml
Taurine	0.1751g
Hypotaurine	0.1091g
Insulin (10mg/ml)	100µl
Penicillin G	0.0120g
Streptomycin sulphate	0.01g

0.0499g of calcium chloride was dissolved into a separate 5ml aliquot of sterile water before being added to the stock solution. Water was added to bring the final volume of the solution to 190ml. The pH of the solution should be 7.4 and the osmolarity should be between 260-280 mOsmol. The solution can be stored at 4ºC for up to two weeks.

Embryo Culture Medium 0.002g of cysteine and 0.08g of bovine serum albumin were added to 20ml of NCSU base medium

Porcine embryos for microinjection were collected for females which had undergone superovulation and insemination. Embryo donors were injected with PMSG six days before the embryo harvest surgery was scheduled and with hCG two days prior to surgery to induce superovulation. Following the hCG injection, the animals were monitored to see whether they went into heat and, if so, how frequently. This informs as to which animals will be used as embryo donors and which will be used as embryo recipients if recipients are required. The day before embryo harvest surgery, donor embryos were inseminated three times (semen sourced from Deerpark

Pedigree Pigs) and were scanned on the day of surgery to determine whether ovulation had occurred and whether any animals had cystic ovaries, which can affect the quality of the embryos harvested. Animals which had ovulated and without cystic ovaries were selected for surgery.

During surgery, the uterus of each pig was checked for the number of ovulation points present and a catheter was then inserted into the uterine horn and the cuff inflated. A bull-nosed needle was inserted into the oviduct and 30mls of NCSU23 HEPES was flushed through the oviduct using a syringe before being collected into a sterile tube. The NCSU23 HEPES used for flushing was then poured into a 90mm Petri dish and any embryos recovered, counted and incubated at 38.5ºC until microinjection.

Genome editors can be delivered into the embryos of many species, such as mice, rats and rabbits, by injection into either the pronucleus or the cytoplasm. Various groups have compared the merits of pronuclear injection against cytoplasmic injection as well as injecting genome editors in different molecular forms (see Horii et al. (2014) as an example). However, it is currently not possible to effectively visualise the pronucleus of ovine and porcine embryos for microinjection. For this reason, all embryo injections carried out in the course of this project were cytoplasmic. All porcine embryos received 2-5pl injections of genome editor RNA into the cytoplasm. The concentrations of genome editors are given in the relevant Results chapter. Injected embryos were cultured in Embryo Culture Medium until blastocyst stage (i.e. 6-7 days).

2.6.2. Preparation of ovine embryos

The following media were used in the preparation of sheep embryos for injection;

Oocyte Wash 4ml of 10x stock of Medium199 was added to 34.5ml of deionised water along with 0.016g of sodium bicarbonate, 0.1192g of HEPES (H-6147) and 0.025g of heparin. On the day of oocyte collection, 435µl of oestrus sheep serum (sourced from the blood of ewes in oestrus) was added to the solution and the pH adjusted to 7.4. The osmolarity was adjusted to fall between 265-290mOsM and the solution was filter sterilised.

Maturation Medium 1ml of 10x stock of Medium199 was added to 8.8ml of deionised water along with 0.021g of sodium bicarbonate and 100µl of 200mM L-Glutamine. The pH was adjusted to 7.6 using either 0.1% hydrochloric acid or 0.1% sodium hydroxide and

the solution's osmolarity adjusted to 292mOsM using deionised water. This stock solution is made fresh each week. On the day of oocyte collection, 2.5ml of oestrus sheep serum, 5µl of 10µg/ml follicle stimulating hormone, 50µl of 1µg/ml lutenising hormone are added to the stock solution prior to filter sterilisation. 5µl of 1µg/ml estradiol is then added.

Stock Solution A The following reagents were mixed together; 3.145g of sodium chloride, 0.267g of potassium chloride, 0.081g of monopotassium phosphate, 0.091g of magnesium sulphate heptahydrate, 0.03g of Penicillin G and 0.3ml of sodium lactate. 49.7ml of deionised water was added to this mixture to give a final volume of 50ml. The solution was then filter sterilised and stored at 4°C. This solution can be stored for up to three months.

Stock Solution B 0.21g of sodium bicarbonate was dissolved in 10ml of deionised water. A small amount of phenol red was then added until the solution took on a slight pink colour. The solution was then filter sterilised and stored at 4°C. This solution is made fresh every week.

Stock Solution C 0.051g of sodium pyruvate was added to 10ml of deionised water before filter sterilisation and storage at 4°C. This solution can be stored for up to two weeks.

Stock Solution D 0.262g of calcium chloride dihydrate was added to 10ml of deionised water before filter sterilisation and storage at 4°C. This solution can be stored for up to three months.

Fertilisation Medium 1ml of Stock Solution A, 1ml of Stock Solution B, 70µl of Stock Solution C and 100µl of Stock Solution D were added to 7.6ml of deionised water. 200µl of oestrus sheep serum was added and the osmolarity of the solution checked to ensure that it was between 265 -285 mOsM. The pH was then adjusted using 0.1M sodium hydroxide until it was between 7.7 and 7.8 before filter sterilisation and storage at 4°C. This solution was made fresh each week.

SOFaaBSA 2ml of Stock Solution A, 2ml of Stock Solution B, 140µl of Stock Solution C and 200µl of Stock Solution D were added to 15ml of deionised water. 400µl of BME amino acids, 200µl MEM, 0.006g of glutamine and 0.08g of BSA were then added to

the solution. The solution were checked to ensure that the pH fell between 7.3 and 7.5 and that the osmolarity fell between 265 -285 mOsM. This solution is made fresh each week.

Ovaries were collected from a abattoir in Bridge of Allan and transported to the lab in Thermos flasks containing PBS warmed to 38.5°C. Prior to collection, all media required for harvest and IVM of the oocytes were placed into an incubator or hot box to warm.

Upon delivery of the collected ovaries to the lab, the temperature of the flasks was checked to ensure that the interior had remained above 26°C. A drop in temperature below this could affect the viability of the oocytes. The ovaries were then rinsed in warm PBS and oocytes aspirated from the follicles using a 10ml syringe with an 18-gauge needle. Every 2-3mls of follicular fluid was added to a tube containing 5-6mls of warmed oocyte wash medium and kept in a water bath until all oocytes had been collected. Oocytes were recovered for the wash medium by emptying the contents of the tube into a Petri dish and removing viable oocytes. Only oocytes with layers of cumulus cells were selected for IVM. Selected oocytes were then subjected to three washes in oocyte wash medium, followed by one wash in maturation medium. 40-50 oocytes were placed into each well of a 4-well plate containing 800µl maturation medium in each well. The plate had been equilibrated in a 5% CO_2 incubator for at least 1 hour prior to use. Oocytes were incubated overnight at 38.5°C and 5% CO_2.

The day after collection, IVF was carried out on the matured oocytes in preparation for injection. As on the day of collection, all media required for IVF was pre-warmed before use. At approximately 22 hours after IVM was begun, the oocytes were washed in oocyte wash medium to remove the cumulus cells and then washed twice in fertilisation medium. Groups of ~40 oocytes were placed into 450µl of fertilisation medium in each well of a 4-well plate and incubated at 38.5°C while the sperm was prepared.

2 hours before IVF was planned, a sperm pellet stored in liquid nitrogen was thawed by agitation in a glass tube in a water bath warmed to 37°C. The semen was then layered at the bottom of a test tube containing 1ml of fertilisation medium and incubated for 2 hours at 38.5°C and 5% CO_2. 25µl of the semen was reserved and pipetted onto a glass slide to check motility of the sample. After 2 hours incubation, 100µl of sperm from the top third of the test tube was removed and 10µl used to perform a cell count. $1x10^6$ sperm cells were added to each well of

matured oocytes and the plate placed back into the incubator for 24 hours before injection.

2.6.3. Injection and transfer of ovine embryos

Prior to injection, embryos were washed in SOFaaBSA to remove any sperm and cumulus cells and sorted to ensure that only live, fertilised embryos were used for injection. As with the porcine embryos, each embryo was injected with a 2-5pl injection of RNA encoding the nuclease in question. All injections were cytoplasmic. RNA concentrations are given in the relevant Results chapter. After injection, embryos were cultured *in vitro* for 6-7 days until they reached blastocyst stage.

Crossbred or Scottish Blackface ewes were used as embryo recipients. All recipients were checked for a body condition of 2.5 or higher and for good health in the udder and teeth. To induce oestrus, progesterone sponges (Chronogest, MSD Animal Health) were inserted into the ewes and were left in the animal for between 11 – 15 days before removal. Ewes normally entered oestrus 36 – 48 hours after removal of the sponge. This was timed so that recipients at 6 days post oestrus were available to receive injected blastocysts at day 6 of development.

Recipient ewes were anaesthetised and prepared for surgery by scrubbing the surgery site with an iodine based solution (Pevidine). A mid line incision was made and the uterus exposed. The number and location (i.e. left or right ovary) of corpus lutea were recorded. Transfers were not carried out if no corpus lutea were found. A blunt needle was used to make a hole in the uterine horn on the same side as the ovary showing corpus lutea and a Drummond pipette containing 2-4 blastocysts was inserted into the hole made by the blunt needle. The incision was then closed and the recipient allowed to recover. Pregnancies were identified by ultrasound scanning at about 28 days post IVF.

2.6.4. *In vitro* analysis of embryos

Injected embryos were cultured for 1 week post injection for *in vitro* analysis. All embryos which had either reached blastocyst stage or had cleaved following injection were collected in 3μl of PBS in individual PCR tubes. Total DNA from each embryo was amplified using the Repli-G Mini kit (Qiagen). Fresh D2 buffer was made from 5μl of 1M DTT and 55μl of Buffer DLB for each batch of amplifications. 3.5μl of Buffer D2 was added to each PCR tube containing an embryo and tubes incubated on ice for 10 minutes to denature the DNA. 3.5μl

of Stop Solution was then added to each tube to halt the denaturation and 40µl of master mix containing DNA Polymerase, Reaction Buffer and nuclease-free water added to each reaction. Reactions were incubated at 30°C for 16 hours followed by an incubation at 65°C for 3 minutes to inactivate the polymerase. The resulting DNA was diluted 1:20 before being used in PCR-based assays. Embryos were analysed for evidence of editing by either a CelI assay (see Section 2.5) or by direct sequencing of PCR products (see Section 2.1.9).

2.6.5. Identification of edited animals

Animals born from embryos injected with designer nucleases were earclipped soon after birth as part of standard husbandry practises. DNA from the earclip tissue was extracted using the DNeasy Blood and Tissue Kit and the region of interest amplified as described in Section 2.5. PCR reactions were then purified and sent for sequencing as detailed in Section 2.1.9. The resulting sequence chromatograms were analysed for evidence of multiple traces at the target site, the presence of which would indicate that the animal had been edited.

3. Establishing vacant developmental niches and blastocyst complementation to create interspecific blood chimeras

3.1. Introduction

3.1.1. Stem cells

Stem cells are primarily characterised by their abilities to self-renew and to differentiate into cells from a range of different lineages. One of the earliest identified stem cells was the haematopoietic stem cell (HSC), which was demonstrated to be capable of repopulating the haematopoietic system of an irradiated host (Till and McCulloch, 1961). HSCs were subsequently found to be one of a number of different adult stem cells, which control the replenishment of cells within a specific organ or tissue. In terms of their ability to differentiate into different cell types, adult stem cells are designated as multipotent, as they generally contribute to a restricted number of related cell types (e.g. HSCs form all cell lineages in the blood but are not considered to make a significant contribution to other tissues).

By contrast, embryonic stem cells (ESCs) are known for their pluripotency as they readily contribute to cell types in all three of the germ layers and, when injected into a recipient embryo, can form a chimeric animal (Gardner, 1968, Bradley et al., 1984) (chimerism is discussed in greater detail in Section 3.1.2). ESCs were first established from mouse blastocyst explants (Evans and Kaufman, 1981, Martin, 1981) and, thus far, have only been successfully isolated from mouse, rat, monkey and human embryos (Buehr et al., 2008, Thomson et al., 1995,

1998). This is mostly due to the fact that ESCs from different species require different culture conditions in order to maintain the delicate balance of transcription factors which sustain the pluripotency of the cells (Wang et al., 2012b), requiring the investment of significant amounts of time and resources to elucidate the specific conditions suited to expanding ESCs from a particular species. This severely restricts the application of this powerful technology (discussed in Section 1.1) in larger livestock species, such as pigs and sheep, where such methods are not currently optimised.

A potential solution to this restriction was developed in the form of induced pluripotent stem cells (iPSCs). By ectopically expressing a number of genes associated with pluripotency in terminally differentiated cells, such as skin fibroblasts, pluripotent cells could be isolated. These cells demonstrated the ability to form teratomas and to contribute to embryos, in the same manner as ESCs (Takahashi and Yamanaka, 2006, Okita et al., 2007). This technique has the potential to allow stem cells to be created for any species, greatly facilitating the production of transgenic animals from species which do not currently have established ESC lines. However, no 'gold standard' iPSC lines have yet been established for any large animal species, largely due to ongoing technical issues with the iPSC method such as the epigenetic 'memory' that iPSCs retain of their original cell type (Barrero and Izpisua Belmonte, 2011, Ohi et al., 2011) and the lack of defined culture conditions as mentioned above.

3.1.2. Chimerism

A chimera is an organism where some cells have a different genotype from others and is generally produced by the injection of donor cells into a recipient animal. The resulting animal will have some cells derived from the recipient, while others will have originated from the injected donor cells. Chimeras can be created using adult animals or embryos. The most common examples of adult chimeras are haematopoietic chimeras, where HSCs are transplanted into an irradiated recipient to reconstitute the haematopoietic niche (Till and McCulloch, 1961). Animals which are chimeric in other organs, such as the liver, can also be produced (Mallet et al., 2002, Wang et al., 2002, Azuma et al., 2007).

Embryonic chimeras can be produced either by injection of pluripotent ESCs or by embryo aggregation. The first rodent chimeras were produced by embryo aggregation (Tarkowski, 1961, Mayer and Fritz, 1974) and were among the first described examples of differing coat

colour in a chimeric animal, where a single animal exhibited both agouti and black fur. Embryo aggregation has also been used to create chimeras in species which do not have naive ESC lines available (Gardner and Munro, 1974, Tucker et al., 1974, Brem et al., 1984, Tachibana et al., 2012). However, this method makes the creation of transgenic animals very complex as a fully transgenic donor embryo must first be produced, making the use of ESCs a more attractive option.

The ability of an ESC to contribute to a chimera is dependent on whether the cell is in a 'naive' or 'primed' state (Nichols and Smith, 2009). Primed stem cells represent a later developmental stage than naive cells and are distinguished by their reliance on Erk signalling and their inability to contribute to chimeras. This is exemplified by mouse ESCs, which are naive, while mouse epiblast stem cells are primed (Tesar et al., 2007). Although the first human and primate ESC lines isolated were in a primed state, the addition of specific factors to the culture conditions has permitted the subsequent derivation of naive human ESCs (Gafni et al., 2013), raising the possibility that naive ESCs from large animal species may eventually be developed.

3.1.2.1. Interspecific chimerism

Interspecific chimeras are produced by using a donor source of cells from a different species to the recipient. As with intraspecific chimeras (i.e. those where both donor and recipient cells are from the same species), interspecific chimeras can be created at the adult and embryonic stages of development. The first published example of an adult interspecific chimera was the transplantation and subsequent growth of human tumour material in immunodeficient mice (Rygaard and Povlsen, 1969). Immunodeficient mice, such as the SCID mouse model can be used to create humanised models of development or disease by transplantation of human cells such as HSCs (McCune et al., 1988, Eyrich et al., 2011), hepatocytes (Mercer et al., 2001) or neurons (Han et al., 2013, Wang et al., 2013c). Transplantation of multiple donor cell types can also produce mice which are chimeric in multiple organs (Wilson et al., 2014). The extensive use of the SCID mouse to produce humanised models, the current deficiencies of the model and future developments have been reviewed by Pearson et al. (2008).

One of the oldest examples of embryonic interspecific chimeras are quail-chick chimeras. These have been used as a method to track cell migration during embryonic development by

transplantation of tissue from one species into the embryo of the other (e.g. transplantation of quail somites into a chick embryo) (Le Douarin, 1973, Le Douarin et al., 1975). Staining the DNA using the Feulgen-Rossenbeck method allows cells from the quail to be easily distinguished from those of the chicken, due to differences in their chromatin structure during interphase. This allows cells originating from the transplant to be tracked as they disperse through the embryo. Chimeras constructed by transplantation of the quail neural crest into chick embryos are viable and neurons from the transplanted quail tissue interact with those from the chick (Kinutani and Le Douarin, 1985). However, these animals develop neurological problems after two months of age, due to immune rejection of the graft. Quail-duck chimeras have also been used in studies of embryonic development (Yamashita and Sohal, 1987).

At around the same time as the creation of quail-chick chimeras, mammalian interspecific chimeras were produced from different mouse species, such as *Mus musculus* and *Mus caroli*, or from rat and mouse cells and embryos. These chimeras were first reported during investigations into early mammalian development and therefore were not taken to term (Gardner and Johnson, 1973, Stern, 1973, Rossant, 1976). However, subsequent efforts only produced viable chimeric animals when different mouse species were used as the donor and host (Rossant and Frels, 1980). This suggests that the evolutionary distance between host and donor species places a restriction on the ability of interspecific chimeras to form, with closely related species more likely to form viable chimeras then those less related to each other.

In larger animals, viable sheep-goat chimeras have been produced by two methods; ICM transfer, where a donor ICM is inserted into a host blastocyst, and blastomere aggregation, where dissociated cells from 4-cell or 8-cell stage embryos are combined (Fehilly et al., 1984). The success of chimera formation was observed to be dependent on the extra-embryonic tissues, such as the trophectoderm and chorionic epithelium, being derived from cells of the same species as the recipient female. Transfer of the ICM from a goat blastocyst to a sheep blastocyst, for example, resulted in the live births of lambs, kids and sheep-goat chimeras from sheep recipients. This need for compatibility between the recipient female and the extra-embryonic tissue could explain the restrictions on the evolutionary distance between donor and recipient seen in the mammalian chimeras described above. Using cells such as blastomeres which can contribute to extra-embryonic tissues could cause rejection of the chimeric embryo by the recipient female if the extra-embryonic tissues derive from the donor species. This is

reinforced by the accidental creation of viable rat-mouse chimeras using ESCs, which do not contribute to the extra-embryonic tissues (Brenin et al., 1997).

As with forming intraspecies chimeras, the success of generating interspecific chimeras also appears to be dependent on the 'primed-ness' of the donor cells in addition to the evolutionary distance between the donor and host species. Attempts to form interspecific chimeras from primed monkey ESCs and mouse embryos failed to produce viable animals and found that although monkey ESCs could become incorporated into the inner cell mass of mouse blastocysts, they do not interact with the mouse host cells or proliferate within the embryo. Other attempts using embryo aggregation and tetraploid complementation either did not create viable embryos or produced embryos which failed to implant *in vivo* (Simerly et al., 2011). This is most likely due to the fact that monkey ESCs are thought to be more primed than those of rodents and are incapable of contributing to chimeras. Chimeric monkeys have thus far only been created as the result of aggregating multiple embryos at the totipotent 4-cell stage (Tachibana et al., 2012). The effect of primed stem cells in chimera formation can also be seen in studies where human ESCs are injected into mouse or chick embryos in studies of early development (Muotri et al., 2005, James et al., 2006, Goldstein et al., 2002). Although all three studies saw evidence of differentiation of the injected human ESCs, Goldstein et al. (2002) and James et al. (2006) observed that the human ESCs tended to remain as a single mass within the embryo and did not disperse through the organism. Muotri et al. (2005) found that their injected ESCs could migrate through the brain of the resulting animals however only 0.1% of all neurons studied were of human origin.

3.1.2.2. Blastocyst complementation

Blastocyst complementation was first reported in 1993 by Chen et al. (1993). They used mice which were deficient for the Rag-2 gene, which is involved in V(D)J recombination during antibody production in B and T lymphocyte maturation (Oettinger et al., 1990). As a result of this, $Rag2^{-/-}$ mice do not have mature T or B cells, but are otherwise healthy and fertile. The lack of mature lymphocytes means that the animals can be considered to have a vacant developmental niche which would have been occupied by B and T cells. By injecting $Rag2^{-/-}$ embryos with WT mouse ESCs, the researchers were able to create mice with mature, functional B and T lymphocytes derived from the injected donor cells. This technique of injecting

cells to fill an empty developmental niche created by mutation or deletion of a gene has been named blastocyst complementation.

Since this original paper, various studies have adapted this technique to correct disease phenotypes (Mueller et al., 2005, Stillwell et al., 2009, Jansson and Larsson, 2010) and to rescue embryonic-lethal phenotypes (Fraidenraich et al., 2004).

3.1.2.3. Blastocyst complementation in interspecific chimeras

One of the most powerful demonstrations of the potential of blastocyst complementation was demonstrated by Kobayashi et al. (2010). In this study, the concept of filling an empty developmental niche with donor stem cells was used with recipient Pdx1$^{-/-}$mice, which do not form a pancreas during embryogenesis and therefore die as neonates. The authors firstly demonstrated that the pancreas niche can be filled by injected mouse ESCs and iPSCs, giving rise to healthy, viable animals with a functional pancreas. They then went on to describe the production of mouse/rat interspecific chimeras through injection of WT rat ES and iPS cells into WT mouse blastocysts and vice versa.

Donor rat cells in these mouse-rat chimeras were shown by GFP expression to contribute to the developing embryo, but not to the placenta. Although the number of live-born chimeras and the average degree of donor cell contribution were considerably lower than those generally seen in intraspecific chimeras, some chimeras did grow to adulthood, with a total development rate of ~20%. These techniques of blastocyst complementation and interspecific chimeras were finally combined by injecting Pdx1$^{-/-}$mouse blastocysts with rat iPSCs. 24% of embryos injected using this protocol resulted in live-born animals. Analysis of the chimeric pancreata showed that the organs secreted pancreatic hormones and were capable of maintaining blood glucose homeostasis in a diabetes model. This is the first recorded production of an entire organ in an animal using blastocyst complementation and has exciting implications for the future of organ transplantation, as the ability to produce species-specific organs for transplants would overcome many of the immunological and physiological difficulties seen in xenotransplantation currently (see Ibrahim et al. (2006) and Yang and Sykes (2007) for reviews).

This work has now been transferred into large animal models, with the creation of donor-cell-derived pancreata in pigs (Matsunari et al., 2013). In this case, pancreas development in a line of transgenic pigs was ablated by the overexpression of Hes1 from the Pdx1 promoter,

leaving the pigs with only vestigial pancreata. To rescue pancreas formation in these embryos, blastomeres from donor pig embryos were injected into morula-stage embryos. The donor embryos were engineered to express orange fluorescent protein, allowing donor-derived cells to be easily identified. Chimeric piglets resulting from these injections were viable, unlike the Hes1 transgenic animals, and were able to control their blood sugar levels. Analysis of a chimera 12 months after birth showed that the pancreas was present and fluoresced orange, indicating that it was formed from the injected donor cells. However, donor-derived orange cells were detected throughout the animal, including in the brain and gonads, areas where the presence of human cells would present an ethical issue. As such, the method still requires significant refinement before human donor cells can be used to produce organs. The progress of this work in large animals is also hampered by the lack of availability of pluripotent stem cells from large animal species, specifically those capable of contributing to chimeras. There are also significant ethical considerations to be taken into account, as described in Section 3.4.4.

Attempts to replicate this success in kidney formation have also run into difficulties (Usui et al., 2012). Work in Sall1$^{-/-}$mice, which die shortly after birth and have malformed or absent kidneys, shows that injected WT mouse ESCs and iPSCs can fill this vacant niche and fulfil the developmental pathways initiated by Sall1. However, of the eight complemented mice pups produced, none survived to adulthood. Injection of rat pluripotent cells has also completely failed to rescue the Sall1$^{-/-}$phenotype, suggesting that some developmental niches are not amenable to being filled by foreign cells, particularly those from a different species. This could potentially restrict not only the types of organs which could be produced by this procedure, but also which donor and recipient species can be used. There are also further issues in using blastocyst complementation to produce organs for human transplant, as any tissues in the organ which do not originate from the empty niche (e.g. blood vessels, neurons and fibroblasts in the pancreas which are not affected by knockout of Pdx1), could still trigger immune rejection of the organ in the eventual transplant recipient.

Blastocyst complementation has also been used to produce ESC-derived thymi in rat-mouse chimeras (Isotani et al., 2011). For these experiments, *nu/nu* mice, which do not have a thymus and therefore do not produce T lymphocytes, were used as recipient blastocysts. Unlike the Pdx1$^{-/-}$ mice used in the work by Kobayashi et al. (2010), the nude mouse phenotype is not

lethal. Injection of the *nu/nu* mice blastocysts with GFP rat ESCs resulted in chimeric animals which had thymi derived from the GFP-expressing rat ESCs and had populations of CD4- and CD8-positive cells in the blood. However, the donor-derived thymi were noticeably smaller in size than would be expected for a mouse thymus and the CD4- and CD8-positive cells were of mouse origin, rather than the donor rat ESCs. When the chimeric thymus was transplanted under the renal capsule of a *nu/nu* rat, CD4- and CD8-positive cell populations were seen in the blood eight weeks after transplantation. This suggests that the injected rat ESCs did not themselves directly contribute to the T cell population, but, by reconstituting the thymus, were able to stimulate the maturation of T lymphocytes from the recipient *nu/nu* mouse and rat cells

3.1.2.4. Blastocyst complementation in the haematopoietic niche

In this project, a two-pronged strategy was used to work towards the development of a chimeric large animal capable of producing human-derived tissues suitable for transplantation, bypassing the technical issues associated with trying to produce a complex, multicellular tissue *in vitro*. Blood was selected as the target tissue as, firstly, there is currently no robust *in vitro* differentiation protocol to produce HSCs from stem cells, although individual lineages can be produced (see Olsen et al. (2006) for review) and, secondly, a range of vacant developmental niches within the haematopoietic system can be produced by different gene knockouts, allowing either whole blood or specific cell populations to be donor-derived. It was decided to focus on Runx1$^{-/-}$ animals, which are incapable of undergoing definitive haematopoiesis during embryogenesis, and therefore do not produce any blood, and Rag1$^{-/-}$ animals, which do not produce B or T lymphocytes. In the first set of experiments, blastocyst complementation was carried out using Runx1$^{-/-}$ or Rag1$^{-/-}$ mice to produce a mouse/rat interspecific chimera with blood derived from the donor rat ESCs (see Figure 3.1). In parallel, genome editors were used in an attempt to create Runx1$^{-/-}$ and Rag1$^{-/-}$large animal models for use in future blastocyst complementation experiments.

3.1.3. Runx1

Runx1 (also known as AML1, Cbfa2 or Pebp2aB) is a transcription factor which is expressed in HSCs and is essential for the initiation of definitive haematopoiesis in the developing embryo

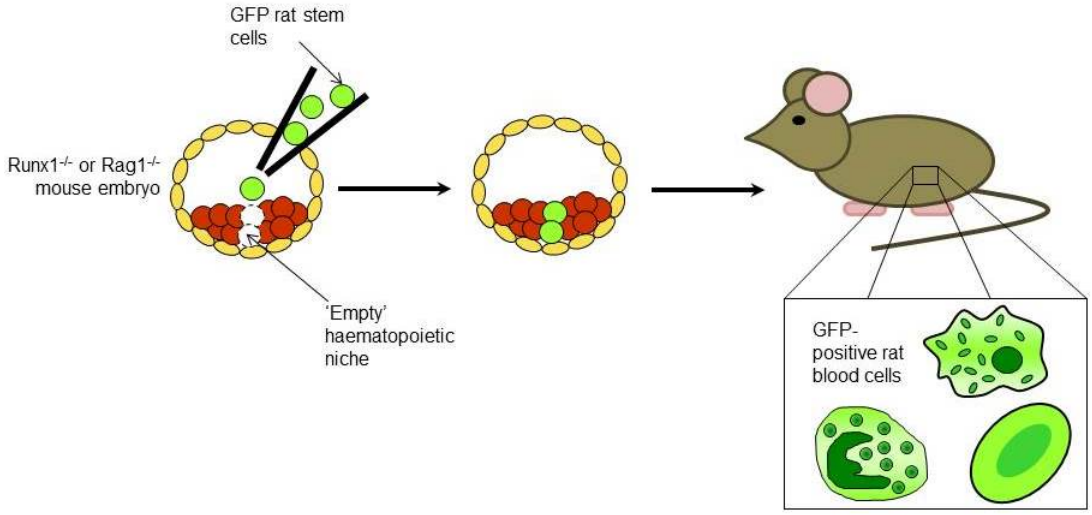


Figure 3.1.: Diagram showing the hypothesised outcome of performing blastocyst complementation experiments using Runx1$^{-/-}$ or Rag1$^{-/-}$ mouse embryos. By injecting GFP-positive rat ESCs, which are capable of forming cells within the haematopoietic lineage, into mouse embryos with a vacant haematopoietic niche, viable chimeric animals with blood cells derived from the rat ESCs should result. In the case of Runx1$^{-/-}$ embryos, all haematopoietic cells should be GFP-positive and, therefore, rat-derived while in the case of Rag1$^{-/-}$ embryos, the B and T lymphocytes populations should be of rat origin.

(Wang et al., 1996). It was first identified as part of the fusion transcription factor Runx1:ETO, which was implicated in the development of acute myeloid leukaemia (AML) (Miyoshi et al., 1991, Erickson et al., 1992). The role of Runx1 in the progression of AML is outside of the scope of this project, but has been recently reviewed by Hatlen et al. (2012).

3.1.3.1. Runx family members

Runx1 is a member of the Runx gene family, the other members of which are Runx2, required for skeletal development (Komori et al., 1997), and Runx3, which is involved in neural development (Levanon et al., 2002). All members of this gene family share a 128 amino-acid long DNA-binding domain with a β-barrel structure (Levanon et al., 2001, Bravo et al., 2001). This domain is homologous to the runt protein in Drosophila (Meyers et al., 1993, Ogawa et al., 1993, Kurokawa et al., 1996) and binds in a sequence-specific manner (Meyers et al., 1993). The function of runt in Drosophila has been reviewed by Canon and Banerjee (2000). The runt homology domain of the Runx proteins also binds the ubiquitously expressed Cbfβ protein to form the heterodimeric transcription factor Cbf. Although Cbfβ does not bind to DNA itself, it

enhances the DNA binding properties of the Runx proteins (Ogawa et al., 1993). In the case of Runx1, this is achieved by relieving the DNA binding inhibition caused by Runx1's negative regulatory region, which lies 3' of the Runt domain (Kanno et al., 1998) and by protecting Runx1 from ubiquitinisation and subsequent degradation (Huang et al., 2001).

Despite their key roles in the development of very different tissues, expression of the Runx proteins is not restricted to cells of a specific lineage, such as blood, as Runx1 is also expressed in neurons and bone (see Section 3.1.3.3 for more details), while Runx3 is expressed in haematopoietic cells (Cai et al., 2011, Wang et al., 2013a). The extent of Runx protein co-expression requires additional regulation to ensure that cell development proceeds towards the correct lineage. To this end, both the bone and neuron development processes are negatively regulated in Runx1-expressing HSCs (Tanaka et al., 2012).

3.1.3.2. Gene and protein structure of Runx1

The Runx1 gene spans 260kb and contains two promoter regions, known as P1 and P2, and 12 exons. P1 is the distal promoter located upstream of the first exon, while P2, the proximal promoter, initiates transcription at exon 3 (Miyoshi et al., 1995, Ghozi et al., 1996). Use of the promoters varies between different stages of haematopoietic development and between different haematopoietic cell populations (Bee et al., 2009, Sroczynska et al., 2009). mRNA transcripts produced from P2 are translated via an IRES rather than by cap-dependent translation, which is used for transcripts from P1 (Pozner et al., 2000). Homozygous deletion of the Runx1 IRES is embryonic lethal. IRES-knockout mouse embryos die from circulation failure at stage E14.5, a later stage of development than Runx1-knockout embryos (as discussed in Section 3.1.3.6), and show reduced numbers of HSCs and increased HSC apoptosis (Nagamachi et al., 2010).

Six of the Runx1 exons are common to all members of the Runx gene family. Exons 2, 3 and 4 form the runt homology domain, which binds a TGTGG consensus site (Wilson et al., 2010) and exons 5 and 6 encode the negative regulatory region, which inhibits DNA binding, and the transactivation domain (Levanon et al., 2001, Huang et al., 2012). Twelve alternatively spliced Runx1 mRNA species have been identified (Levanon et al., 2001) and three of these isoforms have been recognised as performing the majority of the functions of Runx1. These isoforms are Runx1a, Runx1b, both of which are transcribed from the P2 promoter, and Runx1c, which

is transcribed from P1 (Miyoshi et al., 1995). It should be noted that the Runx1a isoform is only found in humans and other primates (Komeno et al., 2014). The 5'UTRs of these three isoforms combine with the IRES to provide a level of translational regulation for Runx1 as 5'UTRs originating from the P1 promoter initiate translation more readily than those from the P2 promoter. The IRES also exhibits different levels of activity in different cell types (Pozner et al., 2000).

3.1.3.3. Expression and function of Runx1 in non-haematopoietic tissues

As mentioned previously, Runx1 expression and function is not restricted to cells of the haematopoietic lineage. Like Runx2, Runx1 also plays a role in bone development. Use of conditional-knockout animals has shown that loss of Runx1 expression after the onset of definitive haematopoiesis affects development of the sternum and may also impact on palette development (Liakhovitskaia et al., 2010). Mice which carry a combination of a Runx1-null allele and a Runx1-hypomorphic allele, and therefore have very limited Runx1 expression levels, show impaired skeletal growth in the time between birth and weaning, pointing to a role for Runx1 in chondrocyte differentiation (Soung et al., 2012).

Runx1 is also expressed in denervated skeletal muscle, where helps to prevent autophagy and muscle wasting after damage to the muscle (Zhu et al., 1994, Wang et al., 2005), and in certain neural populations (Stifani et al., 2008). It is required to stimulate the expression of nociceptive receptors in neurons and conditional knockout mutants show a delayed response to noxious heat and cold stimuli (Chen et al., 2006).

3.1.3.4. Expression and function of Runx1 in haematopoietic stem cells

Runx1 expression is essential for the successful initiation of definitive haematopoiesis in the developing embryo (Wang et al., 1996, Okuda et al., 1996). Definitive haematopoiesis begins on the ventral surface of the embryonic dorsal aorta (Mueller et al., 1994, Medvinsky and Dzierzak, 1996, North et al., 2002, Eilken et al., 2009) and establishes the long-term haematopoietic system of an organism, as opposed to primitive haematopoiesis, which occurs in the extraembryonic tissues and provides only limited, short-term support for the embryo. In mouse embryos, definitive haematopoiesis begins at E10 (Medvinsky and Dzierzak, 1996). Runx1 is required for the survival of cells undergoing endothelial to haematopoietic transition (EHT) in

the dorsal aorta (North et al., 1999, Kissa and Herbomel, 2010, Boisset et al., 2010, Liakhovitskaia et al., 2014). Cells which undergo EHT derive from a subset of cells in the dorsal aorta endothelium known as haemogenic endothelial cells, which are located in discrete intra-aortic clusters (North et al., 2002, Taoudi et al., 2008, Lancrin et al., 2009). These cells are typified by their expression of Runx1, which is induced by signalling through Notch and the FGF pathways from the surrounding endothelial cells via Hes1, which interacts with the transactivation domain of Runx1 (Nakagawa et al., 2006, Butler et al., 2010, Clements et al., 2011, Guiu et al., 2013, Lee et al., 2014). It is thought that expression of Runx1 in these cells in increased by exposure of the cells to fluid sheer stress caused by the initiation of blood circulation in the embryo (Adamo et al., 2009). Although haemogenic endothelial cells express Runx1, it is not necessary for their specification (Sakai et al., 2009, Liakhovitskaia et al., 2014), but is key for the process of EHT.

EHT begins when cells in the ventral endothelium of the dorsal aorta undergo a series of contractions, causing individual endothelial cells lining the vessel to bud off into the circulation and simultaneously take on a haematopoietic phenotype (Kissa and Herbomel, 2010, Boisset et al., 2010). These cells, which express both Runx1b and Runx1c (Challen and Goodell, 2010), go on to become definitive HSCs and populate the haematopoietic niche. Runx1 facilitates this transition by upregulating genes involved in cell migration and adhesion, such as those involved in integrin signalling (Lie-A-Ling et al., 2014). Lack of Runx1 makes these budding events much rarer and any cells which are released from the dorsal aorta apoptose upon leaving the endothelial layer (Kissa and Herbomel, 2010, Boisset et al., 2010). However, once these HSC progenitors have successfully formed, Runx1 expression is again no longer essential (Chen et al., 2009b). The HSC progenitors then migrate firstly to the foetal liver and subsequently to the bone marrow.

This transition of cells from a endothelial phenotype to a haematopoietic one is caused by transcriptional activity controlled by Runx1, which upregulates expression of haematopoietic genes such as CD41 and CD34 (Levantini et al., 2011, Sroczynska et al., 2009, Tanaka et al., 2012) and downregulates endothelial gene transcription (Sakai et al., 2009, Tanaka et al., 2012). This is partly executed by targeting of Runx1 to the transcriptional repressors GFI1 and GFI1b (Lancrin et al., 2012).

Once a HSC has formed, Runx1 takes on a more regulatory role in HSC maintenance, by

slowing proliferation and decreasing apoptosis (Cai et al., 2011), and haematopoietic cell differentiation. This requires Runx1 to co-operate with many other transcription factors to regulate gene expression including SCL, LYL1, LMO2, FLI-1, ERG and GATA2. Transcription can be initiated by a complex containing all or a subset of these factors (Wilson et al., 2010) and many genes controlled by Runx1, including Runx1 itself, are also controlled by the SCL/Gata2/Fli1 triad (Tanaka et al., 2012). Indeed, expression of Runx1 in haemogenic endothelial cells is dependent on binding of the triad to the Runx1 +23 enhancer (Nottingham et al., 2007).

Runx1 helps to regulate HSC differentiation and expansion by regulating the expression of genes such as HMGA2 (Lam et al., 2014), PU.1 and c-MYC. The repression of c-MYC expression by Runx1 may indicate the cell's progressive loss of potency as it differentiates into a HSC and subsequently into specialised cell types (Jacobs et al., 2013). Runx1 also induces global changes in gene expression through epigenetic changes, such as histone acetylation, to initiate haematopoietic development pathways (Lichtinger et al., 2012). This occurs via interactions with epigenetic regulators such as Pcgf1, which ubiquitinates histone H2A and negatively regulates haematopoietic progenitor self-renewal (Ross et al., 2012), and the chromatin remodeling SWI/SNF complex (Bakshi et al., 2010). Runx1 expression is regulated by FGF signalling, which switches to a negative regulatory role after the initiation of Runx1 expression (Pouget et al., 2014) and through a rheostat mechanism with Smad6, which targets Runx1 for proteosomal degradation, thereby regulating the Runx1 dosage in cells (Knezevic et al., 2011).

3.1.3.5. Expression and function of Runx1 in differentiated haematopoietic cells

Runx1 is also active in the development of specialised haematopoietic cell types and it has been suggested that different Runx1 isoforms could determine the differentiation pathways followed by an individual HSC (Komeno et al., 2014). Runx1 stimulates megakaryocyte differentiation from bipotential megakaryocyte/erythrocyte precursors (Tijssen et al., 2011) by repressing the expression of genes such as MYH10 (Antony-Debre et al., 2012) and other genes associated with erythroid development (van Riel et al., 2012). Runx1 is also needed for B-lymphocyte progenitor survival and progression of B cell development (Niebuhr et al., 2013), maintaining T-lymphocyte and platelet populations by regulation of PU.1 (Huang et al., 2008) and in con-

trolling the expression of T-cell receptors (TCRs) (Cieslak et al., 2014) and co-receptors, such as CD4 (Taniuchi et al., 2002).

3.1.3.6. Runx1-knockout phenotype

The importance of Runx1 in the establishment of definitive haematopoiesis was first identified by the creation of the Runx1-knockout mouse (Wang et al., 1996, Okuda et al., 1996). Homozygous disruption of the Runx1 gene causes embryos to die between stages E11.5 and 12.5 due to a lack of definitive haematopoiesis. At stage E10.5, when definitive haematopoiesis normally begins, $Runx1^{-/-}$ embryos show signs of necrosis in the CNS. At the point of embryo death at E11.5-12.5, $Runx1^{-/-}$ embryos exhibit extensive haemorrhaging, which is mostly associated with the CNS and considered secondary to the actual cause of death, and a pale liver, indicative of a block in foetal liver haematopoiesis (Wang et al., 1996, Okuda et al., 1996). Interestingly, this phenotype is recapitulated both in $Cbf\beta^{-/-}$ embryos, demonstrating that both parts of the CBF dimer are required for definitive haematopoiesis (Sasaki et al., 1996), and in embryos where Runx1 is only selectively deleted in cells expressing the endothelial marker Tek (Li et al., 2006). Conversely, restoration of Runx1 expression in only Tek-expressing cells is capable of partially rescuing the $Runx1^{-/-}$phenotype (Liakhovitskaia et al., 2009). $Runx1^{-/-}$ embryos also show a loss of expression of key haematopoietic genes which lie downstream of Runx1 in the development pathway, such as PU.1, M-CSF and c-Myb (Okada et al., 1998).

By contrast, $Runx1^{+/-}$ embryos are viable, although the embryos undergo definitive haematopoiesis earlier and produce fewer haematopoietic progenitors than their wildtype counterparts (Cai et al., 2000). This viability of $Runx1^{+/-}$ is likely to be at least partially due to the ability of other factors such as Runx3 and Gata2 to compensate for a decreased Runx1 dosage (Staber et al., 2014), exemplified by the fact that $Runx1^{+/-}$:$Gata2^{+/-}$ mice are not viable (Wilson et al., 2010).

The creation of conditional deletion mutants in Runx1 has allowed the function of Runx1 in adult animals to be studied in greater detail (Samokhvalov et al., 2006). Loss of Runx1 in adult animals does not cause any severe phenotypic effects but is rather characterised by alterations in different haematopoietic populations, including a large drop in platelet numbers, decreased lymphocyte numbers and an increase in myeloid cells (Growney et al., 2005). This can be linked to the decreased expression of genes such as Cebpa, as lowered Cebpa levels

cause an increased number of progenitors to enter monopoeisis and a decrease in those entering granulopoiesis (Guo et al., 2012) and, as such, reflects the role of Runx1 in specialised haematopoietic lineages (see Section 3.1.3.5), rather than in HSCs.

Deletion of Runx1 in adult animals also increases the number of quiescent HSCs (Ichikawa et al., 2008), indicating that Runx1 is a negative regulator of quiescence. This regulation of quiescence can be linked to the regulation of Runx1 expression over the course of the cell cycle. Runx1 expression increases as cells progress from G1 to S phase, while cells in G1 arrest have lower expression levels of Runx1. This suggests a role for Runx1 in driving cell cycle progression (Bernardin-Fried et al., 2004, Peterson et al., 2005).

3.1.4. Rag1

Rag1 is a protein which, along with Rag2, forms the Rag recombinase complex, found in the immune systems of jawed vertebrates (Leu and Schatz, 1995, Kapitonov and Jurka, 2005). The Rag recombinase stimulates V(D)J recombination of antigen receptor loci, such as those encoding immunoglobulins and T-cell receptors, in developing B- and T-lymphocytes (Schatz et al., 1989, Mombaerts et al., 1992). Animals which are deficient for either Rag1 or Rag2 do not produce mature B- or T-lymphocytes, but are otherwise healthy (Mombaerts et al., 1992).

3.1.4.1. Structure of Rag1

The Rag1 gene consists of two exons, where the large second exon encodes the majority of the Rag1 protein (Mombaerts et al., 1992). Transcription and translation of Rag1 produces a protein 1040 amino acids in length. Early experiments which analysed the activity of truncated Rag1 mutants identified a 'core' region spanning amino acids 384-1010 which, on its own, has the same level of recombination activity as the full-length protein (Sadofsky et al., 1993, Silver et al., 1993). This core region contains a nonamer-binding domain, which binds the nonamer of RSS sites via a GGRPR motif and also plays a role in the synapsis of two RSS sites (Yin et al., 2009). A DDE motif required for DNA cleavage is also found in the core region (Fugmann et al., 2000b).

It is believed that the Rag proteins and RSS sites evolved from an ancient transposon system, as the RSSs bare some homology to the terminal inverted repeats seen in transposons, while the Rag1 and Rag2 proteins share 35-38% identity with the Transib transposase. Despite this

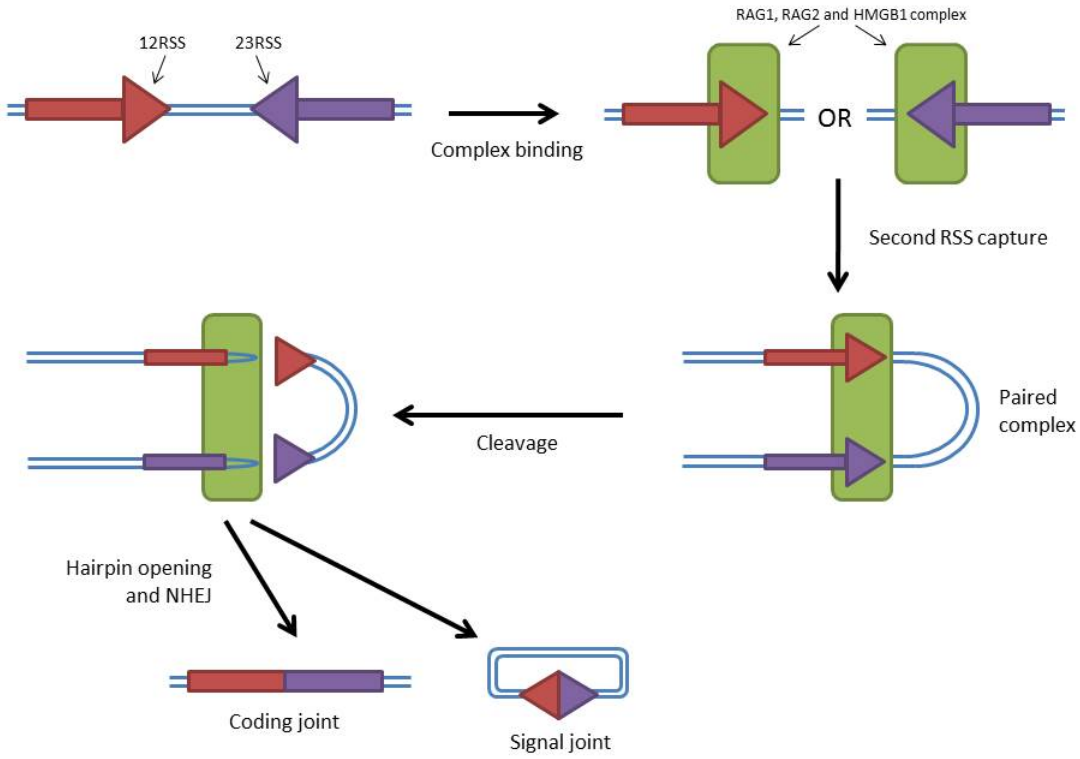


Figure 3.2.: Diagram showing the main stages of V(D)J recombination between two coding sequences as mediated by the Rag recombinase. Note that recombination can be initiated by Rag1 binding to either the 12 RSS or the 23RSS. Figure adapted from Schatz and Ji (2011).

relatively low degree of homology, all residues known to be essential for V(D)J recombination are conserved between Transib and the Rag proteins (Kapitonov and Jurka, 2005, Fugmann et al., 2000b).

3.1.4.2. The Rag recombinase in V(D)J recombination

Although Rag1 is capable of directing V(D)J recombination on its own, complexing with Rag2 vastly increases the recombination efficiency (Oettinger et al., 1990). Figure 3.2 shows the stages of V(D)J recombination.

Rag1 and Rag2 are expressed in B- and T-lymphocyte precursors (Schatz et al., 1989). This limited expression reduces the risk of the Rag recombinase inducing DSBs in the DNA of other cell lineages. Further to this, Rag2 expression levels vary throughout the cell cycle, with the highest levels seen during the G0 and G1 phases, before degradation of Rag2 occurs at the

G1/S boundary (Lin and Desiderio, 1994). This additional level of expression control ensures that the Rag recombinase is not active during DNA replication. The nuclear position of antigen receptor loci forms another control mechanism and prevents the recombination and expression of T-cell receptor loci in B-lymphocytes and vice versa (Kosak et al., 2002).

V(D)J recombination requires the presence of RSSs in the genome. RSSs have a conserved heptamer sequence of CACAGTG, separated by a spacer of either 12bp or 23bp from a conserved nonamer sequence of ACAAAAACC. While the nonamer sequence is essential for V(D)J recombination, the heptamer is not, although it does enhance the binding of Rag proteins to the RSS (Difilippantonio et al., 1996). The difference in the spacer sequence length allows RSSs to be designated as a 12RSS or a 23RSS. Recombination at these loci follows the 12/23 rule, as recombination can only occur between a 12RSS and a 23RSS and not between two RSSs with the same spacer length (Hesse et al., 1989). The 12/23 rule is in part enforced by the Rag recombinase as the Rag1/Rag2 complex make recombination between two 23RSSs energetically unfavourable, while Hmgb1, another protein recruited to RSSs by Rag1, prevents recombination between two 12RSSs (Swanson, 2002).

In addition to the 12/23 rule, V(D)J recombination proceeds in a defined order, with D to J recombination occurring before V to DJ. This is only applicable to heavy chains of antigen receptor loci, as light chains only undergo V to J recombination. It is believed that this precise order is controlled by DNA-binding proteins such as c-Fos, which enhance the binding of Rag1 to the appropriate RSS sites (Wang et al., 2008). There is also epigenetic regulation of V(D)J recombination (Stanhope-Baker et al., 1996), where histone acetylation and chromatin remodeling by the SWI/SNF complex increase V(D)J recombination at certain sites (Kwon et al., 2000). The histone modification H3K4me3 also encourages V(D)J recombination via interactions with the PHD finger of Rag2 (West et al., 2005). This allows hairpin formation at the coding ends, which would otherwise be blocked by the C-terminal domains of Rag1 and Rag2, to occur (Grundy et al., 2010). Although Rag2 recognises H3K4me3 sites throughout the genome, specificity to antigen receptor loci is provided by the binding of Rag1 to the RSS (Matthews et al., 2007, Ji et al., 2010). A more detailed review of how V(D)J recombination is regulated by chromatin structure is provided by Subrahmanyam and Sen (2010).

At the onset of V(D)J recombination, Rag1 binds to a nonamer sequence at an RSS. This binding can be mediated by transcription factors such as Runx1 and Pax5 (Cieslak et al., 2014,

Zhang et al., 2006). Rag1 then recruits Rag2 and Hmgb1 to the RSS (Difilippantonio et al., 1996, Little et al., 2013). While Rag1 and Rag2 together are sufficient to induce DSBs in short stretches of DNA *in vitro*, the DNA-bending protein Hmgb1 is required for successful DNA cleavage *in vivo* (Stanhope-Baker et al., 1996). Hmgb1 is important for mediating DNA bending at the 23RSS to bring the nonamer and heptamer sequences together in the Rag recombinase complex (Ciubotaru et al., 2013, 2014).

This recruitment of proteins to an RSS results in multiple Rag proteins binding to a single RSS, in the form of a Rag1 dimer and a Rag2 monomer (Swanson, 2002). Upon formation of the paired complex (see Figure 3.2), 'recombination centres' are formed, where many Rag1, Rag2, Hmgb1 and potentially other proteins participate in V(D)J recombination (Ji et al., 2010, Schatz and Ji, 2011). Rag1 induces a single strand nick at an RSS between the end of the coding sequence and the first nucleotide of the heptamer (Fugmann et al., 2000a). The freed 3' hydroxyl group then invades the uncut strand via transesterification, causing a DSB and forming a hairpin at the coding end and a blunt ended signal end (McBlane et al., 1995).

The newly formed hairpins at the coding ends of the break are stabilised by protein kinases such as ATM and resolved by NHEJ repair (see Section 1.2.1) (Bredemeyer et al., 2006, Corneo et al., 2007, Deriano et al., 2009), with the hairpins being cleaved by the DNA-PK:Artemis complex (Ma et al., 2002). While the coding ends undergo DNA repair, Rag1 and Rag2 form a post-cleavage complex with Hmgb1 and DNA-PK on the signal joint, possibly to prevent reintegration of the excised DNA (Agrawal and Schatz, 1997). The imprecise nature of NHEJ repair further increases antigen receptor diversity by introducing mutations into the coding sequences (Fugmann et al., 2000a).

Although Rag1 and Rag2 are primarily known for their role in producing antigen receptors in acquired immunity, there is evidence that they also play a role in maintaining cell populations in the innate immune system. The expression of both Rag1 and Rag2 in NK cells increases the cells' genomic stability, allowing cells to exist in a steady state rather than entering apoptosis and creates heterogeneous NK cell populations (Karo et al., 2014). It is possible that further study of other cell populations may uncover a wider range of roles for Rag1 and Rag2.

3.2. Additional Methods

3.2.1. Culture of stem cells

$KMO^{-/-}$ and GFP-Cre mouse ES cells were provided ready for injection by Mr Damian Mole and Dr Peter Hohenstein respectively. Kynurenine 3-monooxygenase (KMO) is an enzyme involved in tryptophan metabolism, however, these cells were used for injections as it was a proven mouse ESC line which would contribute black coat colour to recipients with white fur. By the same token, the GFP-Cre line was also used because it had been previously shown to contribute to chimeras and expressed GFP, which would have simplified identification of chimeras.

DAK31 WT and GFP rat ESCs and OF1 feeders were provided by Dr Tom Burdon. The WT cells have been previously proven to contribute to chimeric offspring (Blair et al., 2012). The GFP line was produced by transfecting WT DAK31 rat ES cells with a pCAGGS-EGFP-IRES-PURO construct and selecting stably transfected, puromycin resistant clones. Transfections and selection of clones was carried out by Dr Stephen Meek.

Prior to irradiation, OF1 feeders were grown in GMEM (Sigma-Aldrich) supplemented with 10% FCS (Sigma-Aldrich), 2mM GlutaMAX (Life Technologies), 1mM sodium pyruvate (Life Technologies), 1:100 dilution of NEAA (Life Technologies) and a 1:500 dilution of 50mM β-mercaptoethanol (Life Technologies). OF1 medium was filter sterilised before use and OF1 cells were grown in gelatinised flasks and passaged using TryPLE Express (Life Technologies). Feeder cells were irradiated with 5 Gy and either directly used for rat ESC culture or frozen in freezing medium (50% FCS, 40% GMEM and 10% DMSO) and stored at -80ºC until required. $3x10^4$ irradiated cells were seeded in each gelatinised well of a 24 well plate for ESC culture.

DAK31 ESCs were cultured in N2B27 media supplemented with LIF and inhibitors. N2B27 media was prepared by combining 100ml of DMEM/F12 with 100ml of Neurobasal media (Life Technologies) and adding 1ml of 100x N2 supplement, 2ml of 50x B27 minus vitamin A supplement and 2ml GlutaMAX (Life Technologies). N2B27 medium was filter sterilised before use. PD0325901 and CHIR99021 inhibitors were purchased from Hilary McLauchlan at the University of Dundee. 1mM aliquots of both inhibitors were made by dilution in N2B27 medium. 10µl of 1mM PD0325901, 30µl of 1mM CHIR99021 and 10µl of rat ESGRO® LIF (Millipore) were added to 10ml of N2B27 medium to make 2i+LIF medium. 2i+LIF medium

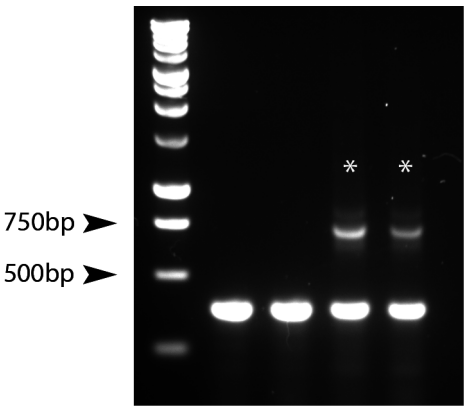


Figure 3.3.: Gel image showing examples of Runx genotyping results. Heterozygotes are indicated by an asterisk.

was stored at 4°C and protected from light.

All cell lines, including feeder cells, used in these experiments were IMPACT-tested by IDEXX Bioresearch before being used for embryo microinjections. No contaminating pathogens were detected in any cells.

3.2.2. Animals and embryo injections

SD rats were provided by Dr Matthew Sharp at the BRR mouse facility at the Royal Infirmary of Edinburgh.

Runx1^{tm1Spe}/J males were sourced from The Jackson Laboratory and mated with WT C57/Bl6 females to establish a Runx1 colony. Individuals were earclipped at weaning as part of standard animal husbandry practises. DNA was extracted from the ear clips using the Fast Tissue-to-PCR kit described in Appendix 3.2.3 and genotyped by PCR using the Runx1 common, wildtype and mutant primers listed in Section A.1. Figure 3.3 shows an example of genotyping results. WT animals produced a single band of 400bp, while heterozygotes have an additional band at 700bp.

Rag1$^{-/-}$ mice were provided from an established colony managed by Dr James Richards. Because the Rag1$^{-/-}$ phenotype is not embryonic lethal, the colony was maintained by mating Rag1$^{-/-}$ animals, ensuring that all offspring would also be Rag1$^{-/-}$. This was validated by flow cytometry, which confirmed that no B or T lymphocytes were present in the blood of the mice (carried out by Dr James Richards).

Female mice of 6-8 weeks of age were selected as embryo donors. Donors were superovulated, mated and the embryos harvested following the schedule below.

Day 1	Donors injected with 5IU of PMSG
Day 2	
Day 3	a.m. - Donors injected with 5IU of hCG p.m. Set up overnight matings
Day 4	Cull donors and harvest 1-cell embryos Culture to 8-cell stage (Day 6)
Day 5	
Day 6	Cull donors and harvest 8-cell embryos Inject 8-cell embryos
Day 7	Cull donors and harvest blastocysts Inject blastocysts Transfer embryos into recipients

Donors were culled by cervical dislocation and embryos injected at different times depending on the stage of embryo required. Micromanipulation and injection of embryos was carried out by Dr William Ritchie.

To harvest 1-cell and 8-cell embryos, the oviducts were dissected out and placed in a 60mm dish with M2 medium (Millipore) and opened by gently tearing the bursa with forceps. Embryos were collected and cultured in microdrops of gassed KSOM medium (Millipore) under mineral oil (Sigma-Aldrich) at 37°C and 5% CO_2 until injection and transfer. Blastocysts were harvested by removing the uterus and flushing each uterine horn with 0.2ml M2 medium before culture in KSOM microdrops.

Injection of ESCs into embryos was carried out using a microinjection rig equipped with the XYClone laser system. A hole was made in the zona pelucida with the laser and embryos were injected with either 2-3 or 5-7 ESCs, in order to find the optimum number for chimerism. Embryos were then cultured in fresh KSOM microdrops until uterine transfer into recipients. F1 recipients were set up by mating with vasectomised males 2 days before transfers were scheduled. In the case of $Rag1^{-/-}$ embryos, recipients and any offspring were housed in isolators to protect the offspring from infection.

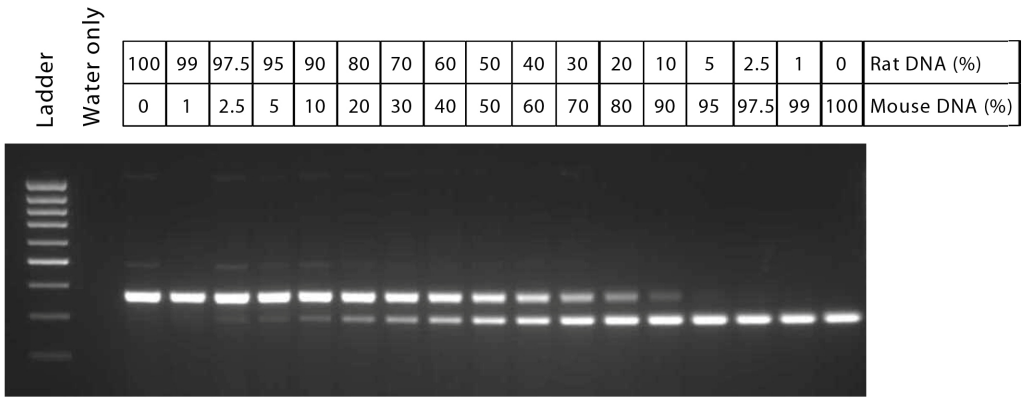


Figure 3.4.: Rex1 PCR on a premixed mouse/rat DNA 'gradient'. The table above the gel shows the ratio of mouse:rat DNA present in each lane.

3.2.3. PCR analysis of animals

Animals resulting from embryo injection of ESCs were culled and tissue samples flash-frozen on dry ice. The Fast Tissue-to-PCR kit (Fermentas) was used to extract DNA from tissues. 2-5mg of tissue was added to a microcentrifuge tube containing 100µl of Tissue Lysis Solution and 10µl of Proteinase K and incubated at room temperature for 10 minutes. Samples were then incubated at 95ºC for 3 minutes before the addition of 100µl of Neutralisation Solution. 4µl of solution from each sample were used in 20µl PCR reactions. Primers against the gene Rex1 were used to distinguish between mouse and rat DNA. The presence of mouse DNA produces a band of 301bp, while a band of 379bp indicates the presence of rat DNA. Figure 3.4 shows the results of carrying out the PCR on a range of samples containing known amounts of mouse and rat DNA.

3.3. Results

3.3.1. Embryo injections to create rat-mouse chimeras

Table 3.1 shows all embryo injections undertaken for the blastocyst complementation experiments.

WT rat and mouse embryos were used to determine whether the production of interspecific chimeras could be established, in addition to the Runx1$^{-/-}$ and Rag1$^{-/-}$ mouse embryos described in Section 4.3.2. Embryos were injected at both the blastocyst stage and at the 8-cell stage as

Donor cells injected	Recipient embryo strain	Recipient embryo stage at harvest	Recipient embryo stage at injection	Number of embryos injected	Time in culture	Embryos transferred to surrogates	Number and species of surrogates	Births
C57/Bl6 mouse ESCs ($KMO^{-/-}$)	WT SD rat	1-cell	8-cell	63	24 hours	25 morula & 27 blastocyst	4 rat	21
DAK31 WT rat ESCs	WT C57/Bl6 mouse	1-cell	8-cell	130	36 hours	126 blastocyst	6 mice	0
DAK31 GFP rat ESCs	$Runx1^{-/-}$ mouse	1-cell	8-cell	63	36 hours	54 blastocyst	3 mice	0
DAK31 WT rat ESCs				66	36 hours	60 blastocyst	4 mice*	N/A
DAK31 WT rat ESCs	$Rag1^{-/-}$ mouse	Blastocyst	Blastocyst	50	~3 hours	64 blastocyst (50 injected and 14 uninjected)	5 mice	24
C57/Bl6 mouse ESCs (GFP-Cre)	$Rag1^{-/-}$ mouse	Blastocyst	Blastocyst	84	~3 hours	54 blastocyst	4 mice	0†
			Total	**456**		**496**	**22**	**45**

Table 3.1.: Summary of embryo injections in blastocyst complementation experiments. * indicates that the recipients were culled at mid-gestation (see Figure 3.6). † indicates that, upon dissection, the surrogates were found to have been pregnant

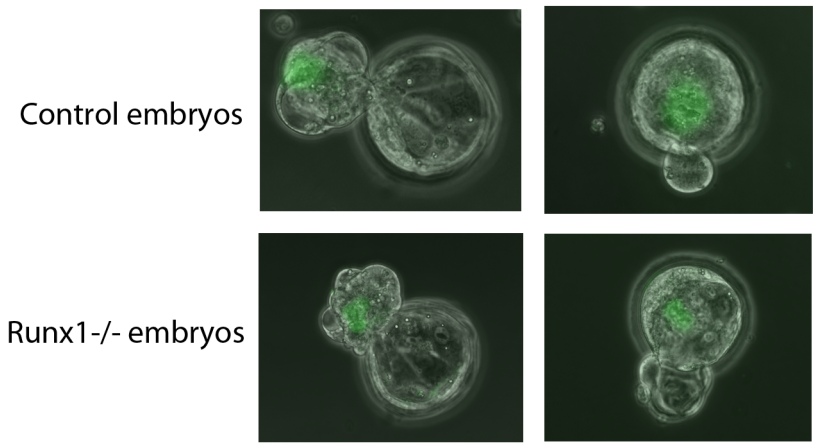


Figure 3.5.: Images of C57/Bl6 control and Runx1-/- mouse embryos injected with DAK31 GFP cells. All embryos shown are at blastocyst stage.

previous work has shown that injecting ESCs into 8-cell stage embryos can result in enhanced contribution from the injected donor cell and, in some cases, an animal completely derived from the donor cells (Poueymirou et al., 2007). Despite numerous rounds of injections and attempts to improve the outcome of these experiments (detailed below), live births were only achieved in injections of rat 8-cell embryos with mouse ESCs and injections of $Rag1^{-/-}$ murine blastocysts with rat ESCs. Animals resulting from these injections were culled and analysed by PCR screening for chimerism, as presented in Sections 3.3.2 and 3.3.3.

After injection of WT rat ESCs into WT mouse blastocysts had failed to produce any live births, DAK31 GFP ESCs were injected into WT and $Runx1^{-/-}$ mouse embryos in order to track the injected cells and make any chimeras easier to detect. Some injected embryos were not transferred but were imaged to determine whether there was any focal localisation of the injected cells within the embryo (Figure 3.5). In all the embryos, a GFP signal from the injected cells could be seen as one group associated with the ICM, suggesting that the cells were localising to the appropriate part of the embryo. However, because no births resulted from this set of injections, the decision was made to use the non-GFP rat ESC line for future injections as, although it was derived from a proven ESC line, the GFP line itself had not yet been proven to contribute to live offspring.

Four recipients of $Runx1^{-/-}$ mouse embryos injected with DAK31 ESCs were culled at mid-gestation to determine the reason behind the lack of live births. Only one implantation site was

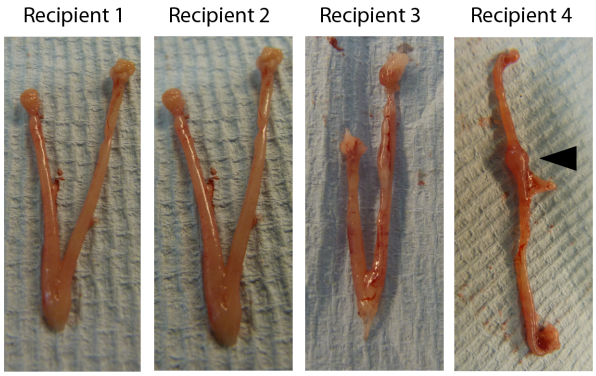


Figure 3.6.: Images of dissected uteri from recipients containing injected Runx1$^{-/-}$embryos culled at mid-gestation. One implantation site was seen in the uterus from Recipient 4 (indicated by arrowhead).

found from all four dissected uteri (Figure 3.6). This strongly suggests that the embryos in this set of injections were being reabsorbed early in gestation. As a result of this, future injections were carried out using blastocysts rather than 8-cell embryos. This was in an effort to reduce the length of time the embryos spent in culture prior to transfer, as the culture conditions were thought to be potentially affecting embryonic development. It was also decided to switch from Runx1$^{-/-}$ host embryos to Rag1$^{-/-}$, to ascertain whether a different source of embryos could solve the problem. In addition, the non-lethal Rag1$^{-/-}$ phenotype should allow for a greater number of live pups to be born. Live births were achieved from injection of Rag1$^{-/-}$ blastocysts with DAK31 WT ESCs and the offspring were analysed (see Section 3.3.3).

The final set of injections, where Rag1$^{-/-}$ embryos were injected with GFP-Cre-ERT2 mouse ESCs, produced no live offspring. Recipients were dissected one week after the litters were due in order to determine the fate of the embryos. Dissection revealed that 3 of the 4 recipients had been pregnant and each animal contained between 2-4 implantation sites. This strongly suggests that all three recipients had given birth to offspring, which were cannibalised soon after birth. It is likely that the cannibalism was triggered by either all the pups being born dead or dying or by the recipient's rejection of the litter due to an external stressor (Weber et al., 2013). Although litter loss is considered a rare phenomenon in mice, rates can vary widely and it can occur in up to 50% of litters (Weber et al., 2013). That being said, the loss of three litters is highly unusual and, combined with the relatively low number of implantation sites found (a total of 8 from 54 transferred blastocysts), suggests that there were problems with the

development of the embryos, causing the majority to be reabsorbed.

3.3.2. Analysis of SD rat/mouse chimeras

Of the 21 animals resulting from injection of SD rat embryos (white fur) with C57/Bl6 mouse ESCs (black fur), one was visually chimeric as it had both white and black fur, indicating contribution from the injected ESCs. All other animals had white coats. The visually chimeric animal died 17 days after birth. Post-mortem examination showed that the animal's kidneys and heart were both enlarged which, combined which other observations, indicate that the animal died from a circulatory disorder. Tissue samples were collected as part of the post-mortem examination and analysed for the presence of mouse DNA by PCR (Figure 3.7). Mouse DNA was clearly detected in the kidney and muscle, confirming that the animal was a rat/mouse chimera. Although the black skin colouring on the chimera was derived from the injected mouse ESCs, no mouse DNA was detected in the PCR assay, despite the fact that at least some of the cells in the sample must have originated from the donor mouse ESCs. This result has likely been affected by a combination of the DNA extraction kit used, which does not completely lyse the tissue sample and therefore only extracts DNA from the edges of the sample which have been lysed in the incubation time, and the somewhat low sensitivity of the PCR assay, as discussed below.

The remaining animals from these injections were culled at weaning and tissue samples taken for analysis (Figure 3.8). No mouse DNA could be detected by PCR in any of the tissues sampled, suggesting that they were not chimeric. However, Figure 3.4 shows that the PCR screen is only capable of detecting a minimum of 2.5% mouse DNA in rat DNA. Although Figure 3.7 shows that this level of sensitivity is adequate to demonstrate tissue chimerism in an overtly chimeric animal, it is likely that a chimera where 1% or less of the total cells are derived from the donor cells will not be picked up by the screen. Therefore, in this and similar assays presented in this chapter, the possibility that extremely low-level chimeras were produced but not detected cannot be discounted.

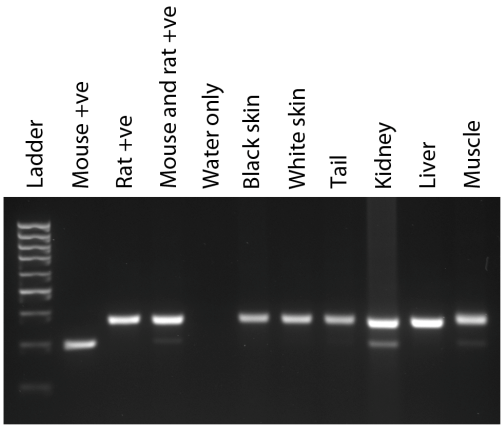


Figure 3.7.: PCR analysis of rat/mouse chimera. Mouse DNA contributed by the injected ES cells can be seen in samples taken from the kidney and muscle

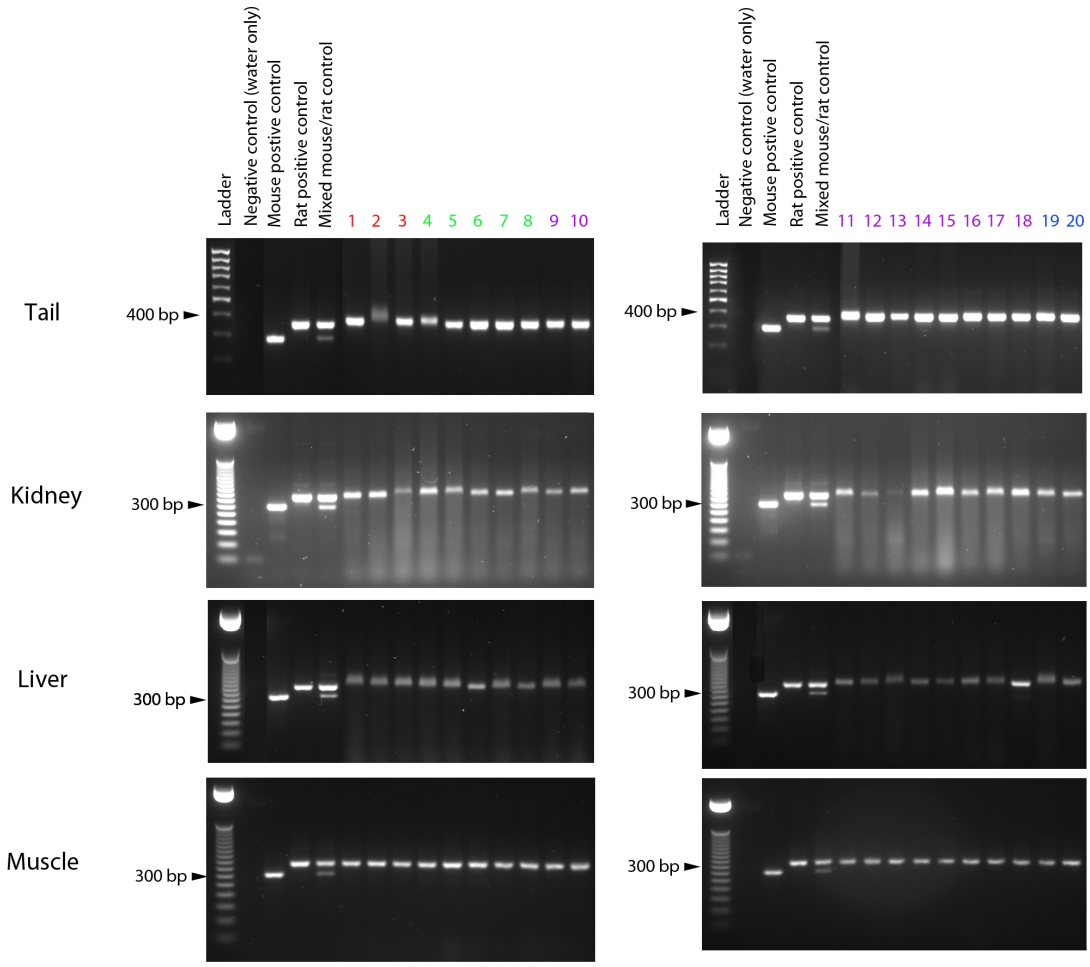


Figure 3.8.: PCR screen for chimerism in remaining animals resulting from injection of mouse ES cells into rat embryos. Numbers with the same colour indicate members of the same litter.

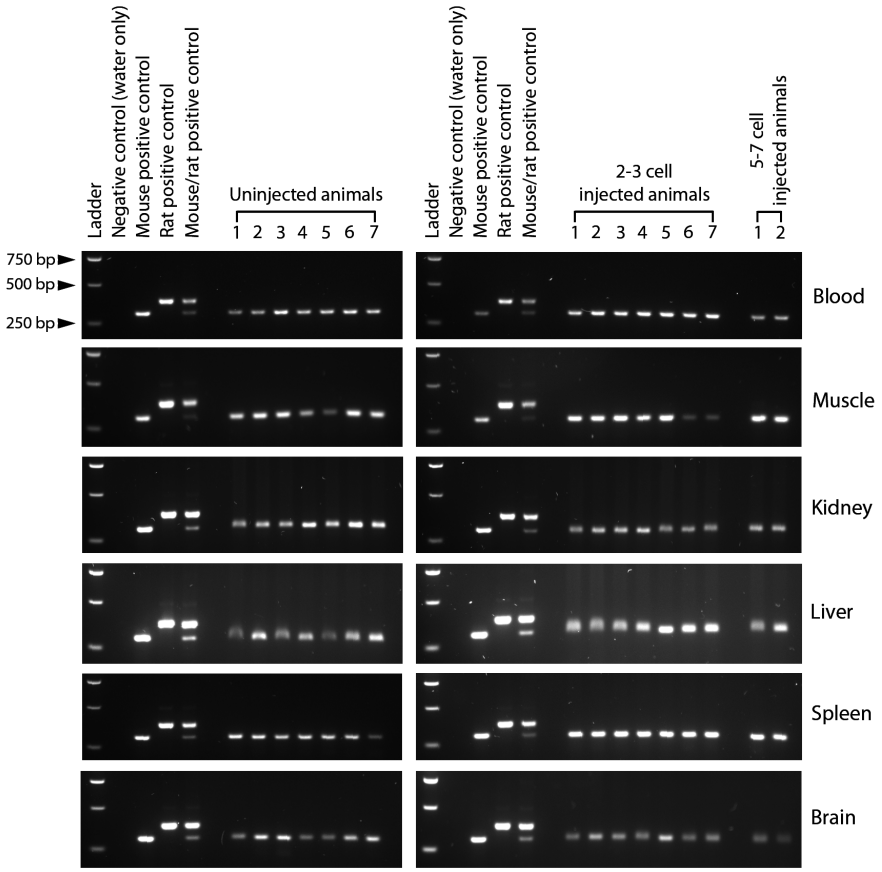


Figure 3.9.: PCR analysis of Rag1$^{-/-}$ mouse embryos injected with rat ESCs. Animals are grouped according to the number of cells injected.

3.3.3. Analysis of animals resulting from injection of rat ESCs into Rag1$^{-/-}$ embryos

Two separate sets of injections involving Rag1$^{-/-}$ embryos produced live-born animals. As the recipient Rag1$^{-/-}$ mouse strain had black fur and the donor DAK31 ESCs would produce agouti fur, identification of chimeric animals may not have been visually apparent. Therefore, as in the previous section, all animals were culled at weaning and tissue samples taken for PCR analysis. The results can be seen in Figures 3.9 and 3.10. No evidence of rat DNA was detected in any of the tissues or animals analysed, despite the use of a proven ESC line. The mouse/rat positive control used in each PCR assay also indicates that the assay should be capable of detecting any significant amounts of rat DNA present in the samples and suggests that the animals are either not chimeric or only possess a very low level (<5%) of chimerism.

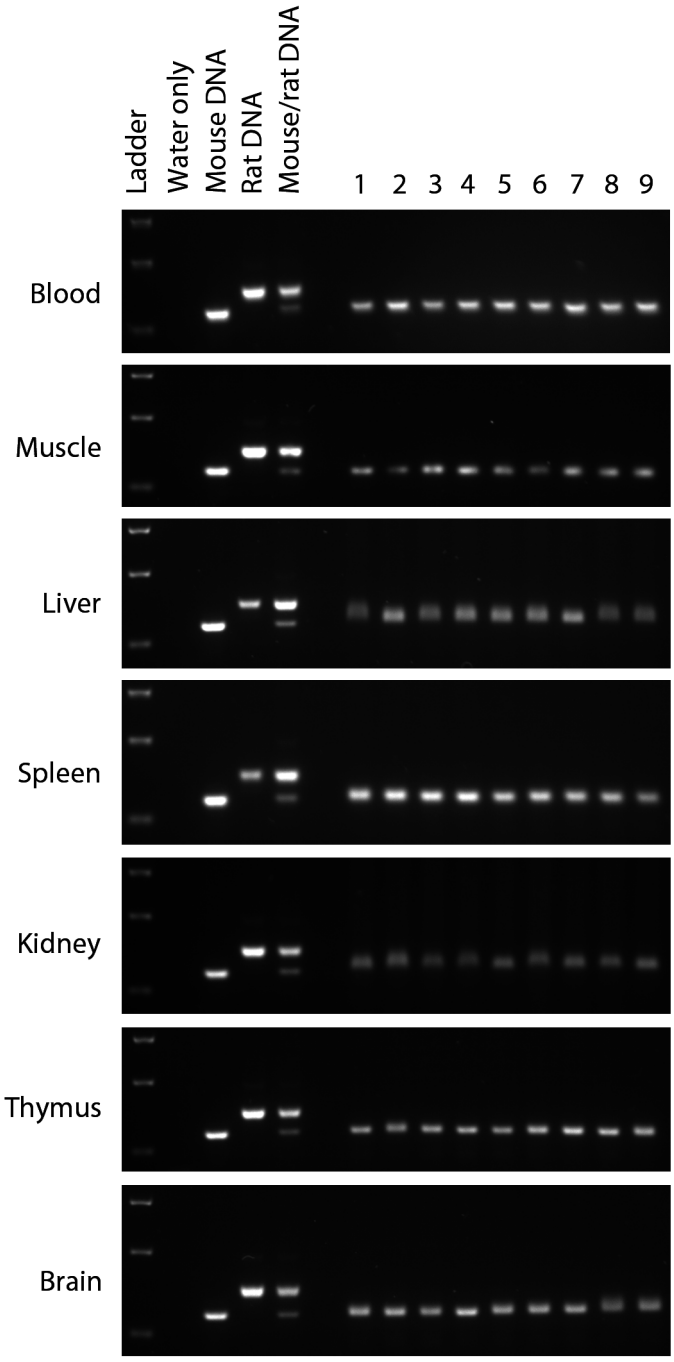


Figure 3.10.: PCR analysis of Rag1$^{-/-}$ mouse embryos injected with rat ESCs. In this set of injections, all embryos were injected with 5-7 cells in an attempt to produce some chimeras.

3.3.4. Runx1 TALENs in large animals

In parallel to the blastocyst complementation experiments, genome editors were designed to reproduce the Runx1$^{-/-}$ and Rag1$^{-/-}$ genotypes in large animals. This would create suitable recipient embryos for future blastocyst complementation work.

Three pairs of TALENs targeted to exon 2 of the porcine Runx1 gene were designed (Figure 3.11), with the aim of creating non-functional Runx1 alleles in the pig. Exon 2 was selected as it is equivalent to exon 3 in the human Runx1 gene, which is the first exon common to both Runx1b and Runx1c transcripts (note that, as previously mentioned, the Runx1a isoform is only expressed in primates). Once constructed and checked by sequencing, as detailed in Section 2.2, each pair of plasmids encoding Runx1 TALENs was co-transfected into PK15 cells with a GFP plasmid to indicate transfection efficiency. DNA from transfected cells was collected three days after transfection and used in a T7 endonuclease assay to check for signs of TALEN activity (Figure 3.12 (a) and (b)). No evidence of TALEN activity was found. To ensure that the Runx1 target locus in PK15 cells was the same as the porcine genome sequence used to design the TALENs and PCR primers, the PCR product amplified by the Runx1 primers was sent for sequencing. Alignment of the sequenced PCR product with the porcine Runx1 sequence showed complete consensus between the two, indicating that the the DNA sequences to which the TALENs were designed were present in PK15 cells (Figure 3.12 (c)).

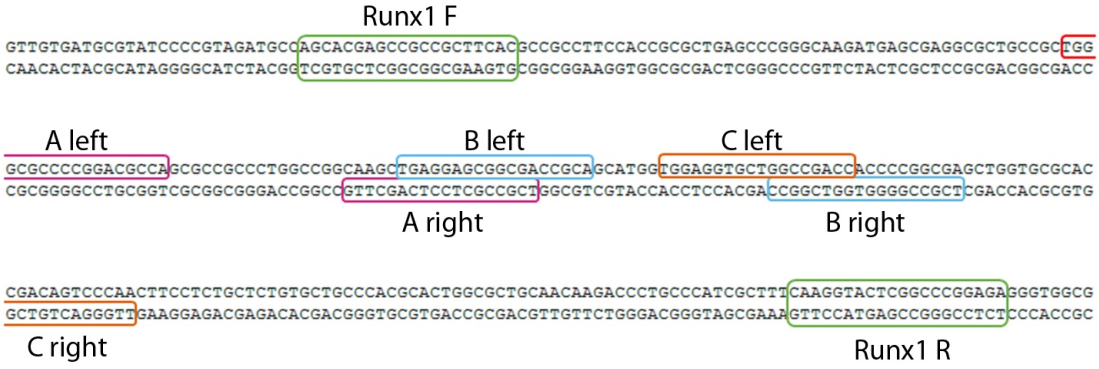


Figure 3.11.: Positions of TALEN binding sites in exon 2 of the porcine Runx1 locus. Primers used for analysis are shown by green boxes.

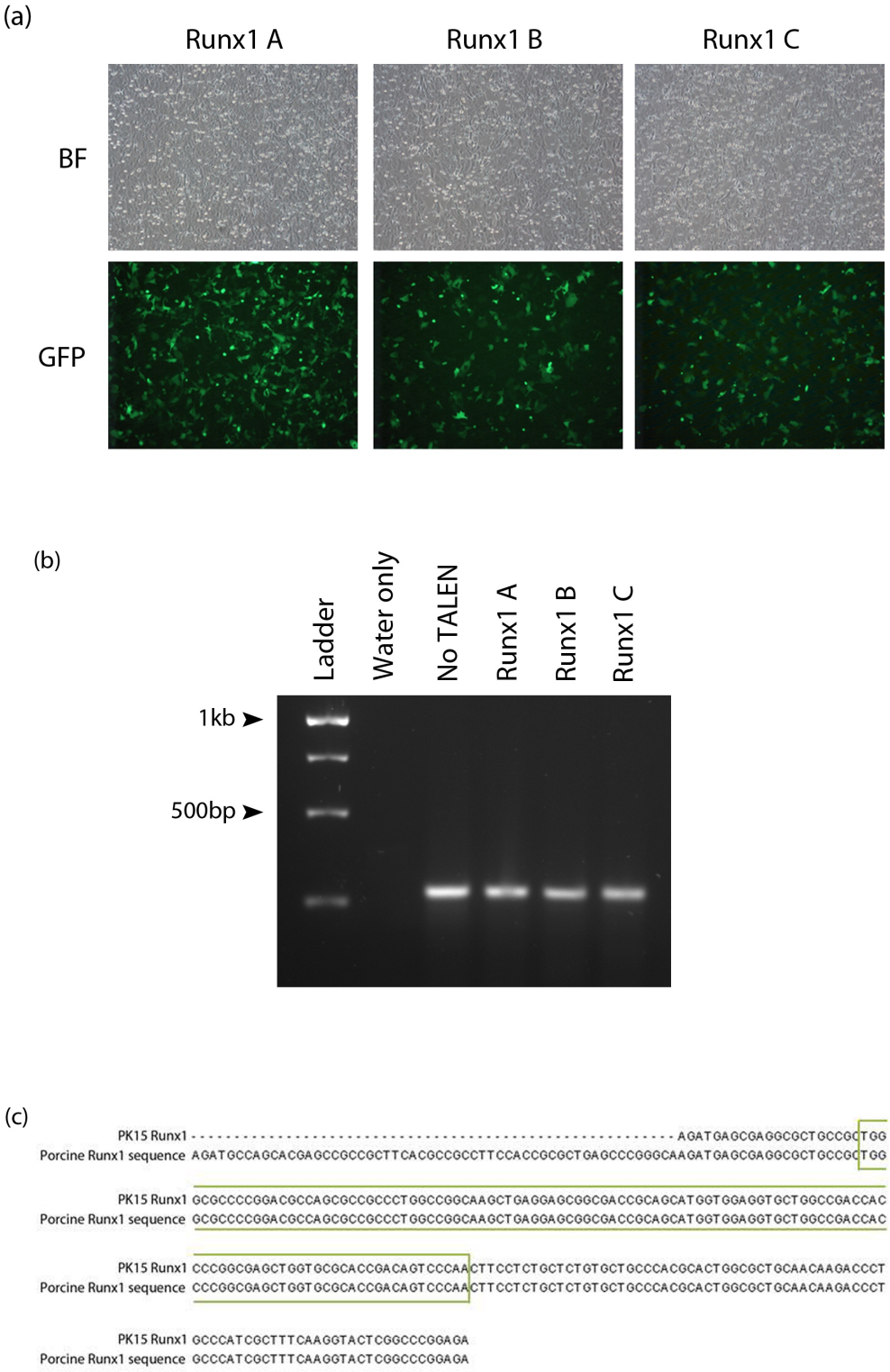


Figure 3.12.: Transfection of PK15 cells with TALENs targeted against Runx1. (a) Brightfield and GFP images of PK15 cells transfected with one of the Runx1 TALEN pairs and a GFP plasmid to show transfection efficiency. (b) Results of T7 assay of DNA isolated from transfected cells. Note the lack of cleavage products indicating the absence of TALEN activity. (c) Sequence alignment of Runx1 TALEN target site sequenced from PK15 DNA aligned against the Runx1 sequence from the porcine genome. The area targeted by the Runx1 TALENs is highlighted in the green box.

3.3.5. Rag TALENs and CRISPR sgRNAs in porcine and ovine cells

In order to develop Rag1$^{-/-}$ pigs and sheep, TALENs were designed against exon 2 of the porcine and ovine Rag1 genome sequences, which contains the majority of the coding sequence of the gene (Figure 3.13 (a) and (b)). sgRNAs for use in the CRISPR/Cas9 system were also designed against the sheep Rag1 locus (Figure 3.13 (c)). Rag1 TALENs were initially constructed into the pRCIScript vector (performed by Mrs Claire Neil) and subsequently reconstructed into the Sangamo ELD/KKR vectors once these plasmids became available. Transfection of Rag1 TALEN pairs as mRNA produced from the RCIScript vector into either PK15 or OEF cells and subsequent CelI assay showed that ovine Rag1 TALEN pair C was active in OEFs (Figure 3.14), albeit with a seemingly low efficiency. In order to visualise the TALEN activity it was necessary to increase the exposure of the original image to the point where the other bands on the image were saturated. Therefore, it is not possible to accurately quantify the amount of editing seen. No TALEN activity was seen in PK15 cells. GFP expression in the transfected cells indicated that the transfection efficiencies were as expected.

A repeat transfection and CelI assay of OEFs with plasmids encoding the Rag1 TALENs did not show any evidence of TALEN activity, including from pair C, which had previously shown editing activity (Figure 3.15 (a)). A number of troubleshooting steps were undertaken to determine the cause of this loss of activity, as it was a problem common to other genome editors used in this project. Firstly, DNA from previously transfected cells was analysed using the T7 endonuclease assay as previous experience with this assay in the lab (and subsequent published data in Vouillot et al. (2015)) suggested that it was more sensitive to the presence of heteroduplexes than the CelI enzyme. This step also helped to address whether there were any problems with the CelI assay reagent, such as a defective enzyme batch. The results of the T7 assay (Figure 3.15 (b)) show that the T7 endonuclease does appear to be more sensitive to mismatches than the CelI enzyme as cleavage products were seen on the gel which were not produced during a CelI assay. Unfortunately, the presence of cleavage products in the 'No TALEN' control lane indicated that these cleavage products were not caused by any editing activity on the part of the TALENs, but rather from the presence of endogenous SNPs in the OEF genome.

To ensure that transfection of the TALEN plasmids into OEFs was successful and not failing

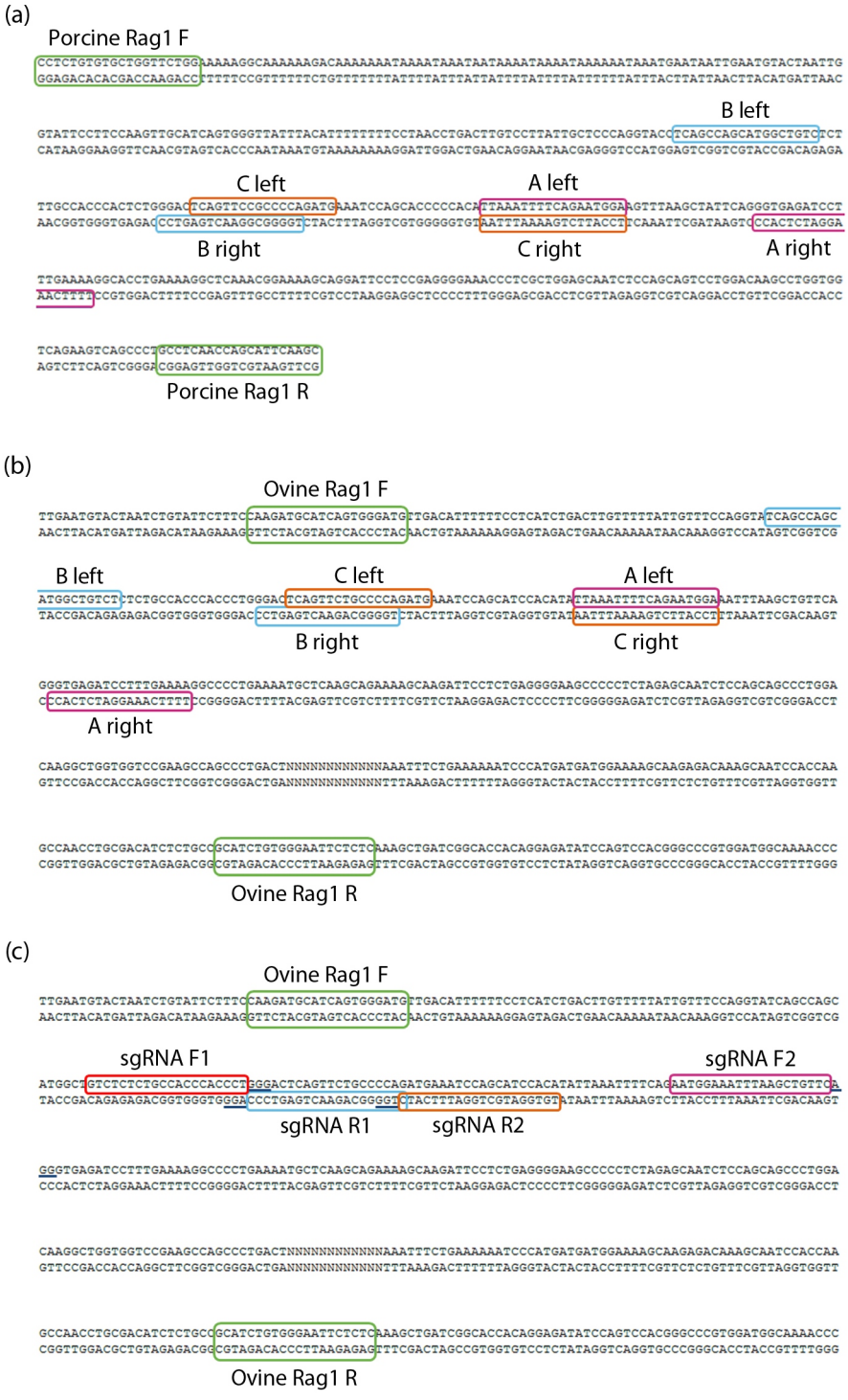


Figure 3.13.: Positions of binding sites of genome editors within exon 2 of the porcine and ovine Rag1 loci. PAMs for sgRNAs are underlined in dark blue. (a) Porcine Rag1 TALENs. (b) Ovine Rag1 TALENs. (c) Ovine Rag1 sgRNAs for CRISPR/Cas9. Primers used for analysis are shown by green boxes.

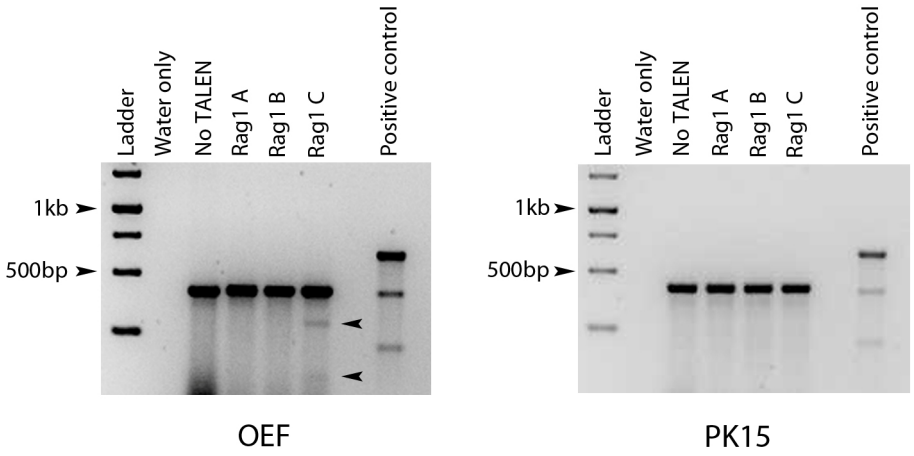


Figure 3.14.: Results of CelI assay on OEFs (left) and PK15 cells (right) transfected with Rag1 TALEN pairs A, B and C. Cleavage of editing products in can be seen in OEFs transfected with Rag1 TALEN pair C (indicated by arrowheads).

due to operator error, two sets of new transfections were carried out. One set was carried out by Dr Chris Proudfoot (Figure 3.15 (c)) while the other involved the co-transfection of Rag1 TALENs with a GFP plasmid to assess transfection efficiency (Figure 3.15 (d) and (e)). CelI assays on DNA from both sets of transfections yielded negative results.

sgRNA oligos were cloned into the pSpCas9(BB) vector and transfected into OEFs (Figure 3.16). Although GFP expression indicated that transfections had been successful, assessment of the transfected DNA in a CelI assay did not show any evidence of CRISPR/Cas9 editing activity.

Finally, it was suggested that the endogenous target sites within the PK15 and OEF Rag1 loci may not match the sequences given in the porcine and ovine genomes due to the presence of SNPs and other point mutations. To investigate whether this was the case, the PCR product amplified using Rag1 primers in both OEFs and PK15 cells was sent off for sequencing and aligned back to the original porcine and ovine sequences used to design the TALENs (Figure 3.17). Some SNPs were found in the PK15 Rag1 target site however, the cells were heterozygous at every SNP position, with one allele being the same nucleotide as the porcine genomic sequence. No differences which would be predicted to have an adverse effect on TALEN and sgRNA binding were detected.

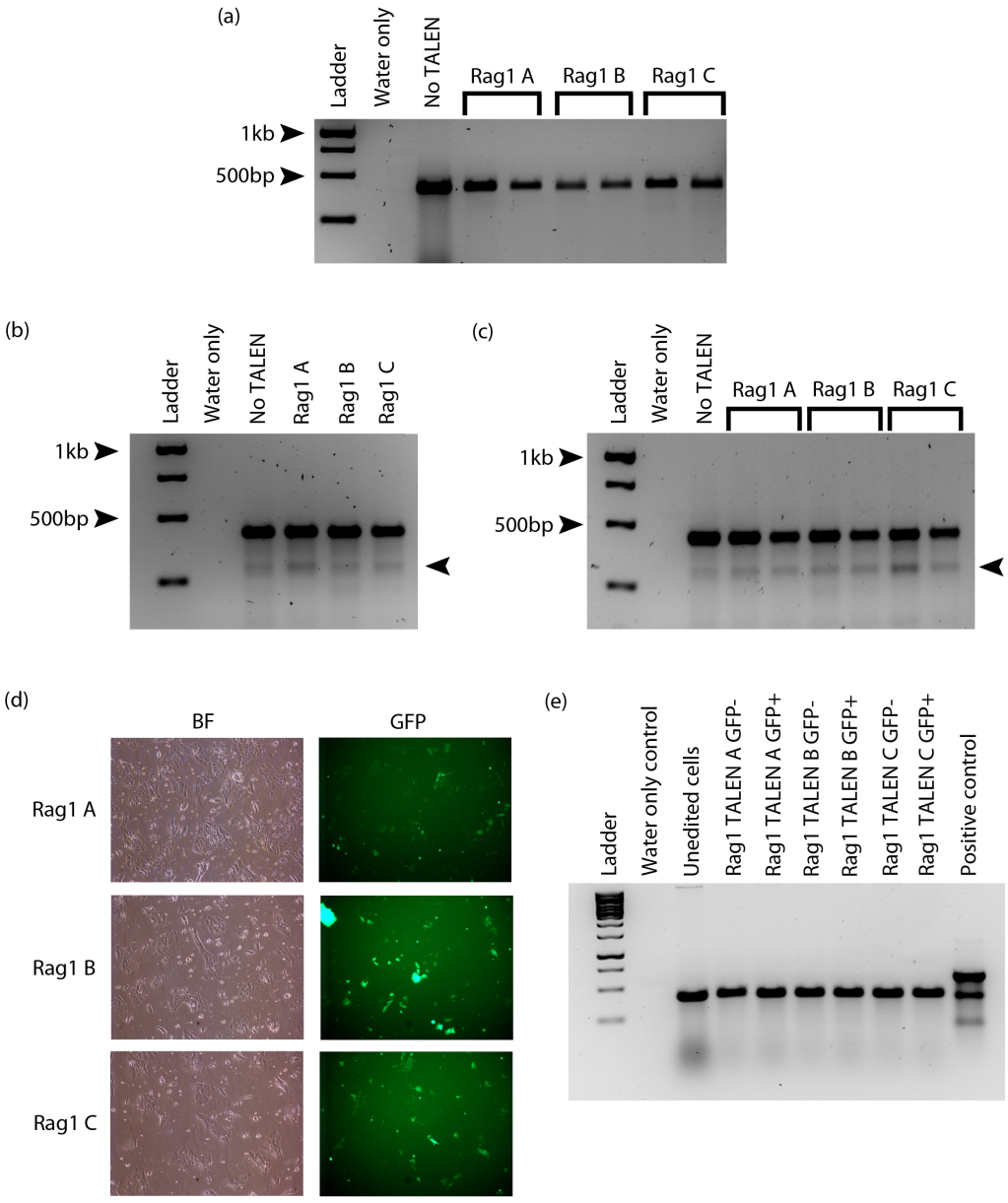


Figure 3.15.: Results of attempts to rectify lack of activity detected using the Rag1 TALENs. (a) CelI assay of duplicate transfections of Rag1 TALEN in Sangamo ELD/KKR vectors. No evidence of editing was detected, including in pair C, which had previously given a positive result in this assay. (b) T7 assay on OEFs transfected with the Rag1 TALENs. Note that cleavage products are visible in all samples, included the no TALEN control, indicating that cutting is not due to the presence of editing events resulting from TALEN activity. (c) T7 assay on transfections carried out by Dr Chris Proudfoot. In this set of transfections, duplicate transfections were carried out for each TALEN pair. As in (b), cleavage products are seen in every lane, including the no TALEN control. (d) Brightfield and GFP images of PK15 cells transfected with one of the Rag TALEN pairs and a GFP plasmid to allow sorting for transfected cells. Cells were sorted by GFP expression and GFP-positive and GFP-negative populations tested for TALEN activity by CelI assay, the results of which are shown in (e). No evidence of TALEN activity was seen in any of the sorted populations.

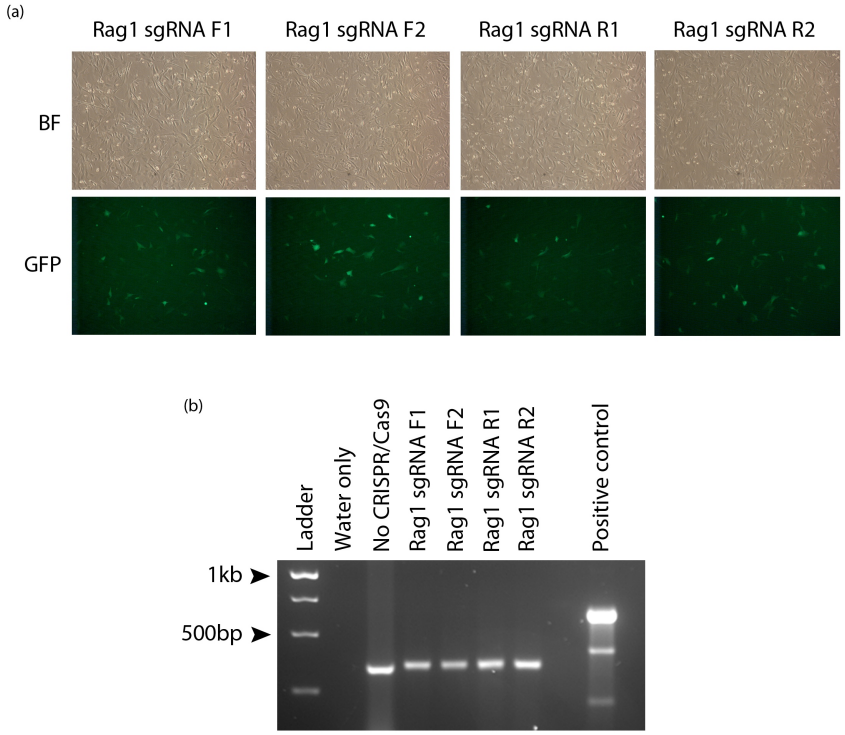


Figure 3.16.: Validation of Rag1 sgRNAs in OEFs. (a) Brightfield and GFP images of cells transfected with sgRNAs in the pSpCas9(BB) vector. Images at 100x magnification. (b) CelI assay on OEFs transfected with one of the Rag1 sgRNAs and Cas9. Note the absence of cleavage products.

Figure 3.17.: Sequence alignments of Rag1 TALEN and CRISPR target sites sequenced from PK15 or OEF DNA aligned against their counterparts in the porcine or ovine genome. Areas targeted by the Rag1 editors are highlighted in green boxes. SNPs detected in the PK15 Rag1 sequence are marked with an asterisk.

3.3.6. Rag TALENs and CRISPR sgRNAs in porcine and ovine embryos

In parallel to the work in cells, Rag1 TALEN pair C mRNA was injected into ovine and porcine embryos. These embryos were harvested from donors but were not subsequently required for their intended microinjection experiments and would otherwise have been disposed of. As harvesting these embryos takes a significant amount of effort and, in the case of sheep, embryos are only available during the few months of the breeding season once a year, it was decided to inject the Rag1 genome editors in these embryos for *in vitro* analysis rather than have the embryos go to waste.

The TALEN RNA was injected at a concentration of 2ng/µl. The resulting blastocysts were assessed *in vitro* for TALEN activity by CelI assay (Table 3.2). The binding sites recognised by the TALENs in pair C are conserved between the porcine and ovine Rag1 sequences so the same TALEN pair can be used in both species. No edited embryos were identified as a result of TALEN mRNA injection. Some injected ovine embryos were reserved for transfer into recipient animals to produce live-born animals (Table 3.3). Six lambs were born as a result of injection of Rag1 TALEN mRNA into embryos. Amplification and sequencing of the Rag1 target site from these lambs only produced a single trace of the wildtype Rag1 sequence on the resulting chromatograms, indicating an absence of editing. These results contrast starkly with editing efficiencies previously obtained in the lab, where 18% of in vitro embryos and 21% of live born piglets resulting from injection of TALEN mRNA were edited (Lillico et al., 2013). However, the TALENs used in this case, which were targeted against the gene RELA, were shown to have a higher activity in cells than the Rag1 C TALEN.

To discover whether multiplexing the Rag1 sgRNAs would improve CRISPR/Cas9 activity, ovine embryos were injected with pairs of the Rag1 sgRNAs. sgRNAs were paired based on the distance between the two cut sites and whether deletion of that distance would cause a frameshift mutation. sgRNAs F1 and R2 were injected in one injection mix, while F2 and R1

Construct	Species	Number of embryos analysed	Number of edited embryos
Rag1 TALEN C (RCIScript)	Sheep	8	0
	Pig	22	0

Table 3.2.: Table summarising embryos injected with genome editors targeted against Rag1 and analysed at blastocyst stage.

Construct	Embryos injected	Cleaved embryos	Embryos transferred	Pregnancies	Lambs	Edited lambs
Rag1 TALEN C (RCIScript)	114	56 (49%)	22 (19%)	2	3	0
Rag1 TALEN C (Sangamo)	43	28 (65%)	12 (28%)	2	3	0
Rag1 F1 and R2 CRISPR	31	16 (52%)	2 (6%)	1*	1	0
Rag1 F2 and R1 CRISPR	30	11 (37%)	4 (13%)			

Table 3.3.: Table summarising microinjections carried out in ovine embryos with genome editors designed to target Rag1. Percentage values for cleavage and blastocyst formation are given in relation to the number of embryos injected. * indicates that both embryos injected with the Rag1 F1 and R2 CRISPRs and embryos injected with the F2 and R1 CRISPRs had been transferred into the recipient which then became pregnant.

were used as a pair in a separate mix (Table 3.3). sgRNA-injected embryos were transferred into recipients and 1 live-born lamb was obtained. DNA from an earclip taken from the lamb was amplified and sent for sequencing however, no edited alleles were detected.

3.4. Discussion

Issues connected to the design and use of genome editors in this project are primarily discussed in detail in Chapter 6. Therefore, this discussion section will focus on the blastocyst complementation experiments.

3.4.1. Difficulties in producing viable interspecific chimeras and potential solutions

Although the blastocyst complementation experiments produced a live-born rat/mouse chimera, it died 17 days after birth from a circulatory disorder. Given that the greatest amount of mouse DNA in the chimera was seen in the kidneys, which were discovered to be enlarged at post-mortem, it is possible that the animal's chimerism contributed to its death. However, Kobayashi et al. (2010) note that the interspecific pancreata generated in their experiments tended to grow to the size normally found within the host animal, rather than the size that would normally be associated with the donor cells (i.e. a rat-derived pancreas in a mouse would grow to a mouse size). In this case of injecting mouse ESCs into a rat blastocyst, it seems highly unlikely that the fact that the donor cells were from a different species would cause the kidney enlargement, especially as, if organ size was determined by the donor cells alone, one would expect the kidneys to be considerably smaller than those in a WT rat. It seems more likely that the enlarged kidneys were caused by a disorder which was not directly caused by the donor cells.

The lack of live births from injections using WT and Runx1$^{-/-}$ mouse embryos suggests a problem with the embryo culture conditions, causing the embryos to be reabsorbed shortly after transfer into the recipients, as demonstrated by Figure 3.6, where only one implantation site was seen in recipients culled at mid-gestation. Considering that implantation should occur at the late blastocyst stage, and therefore shortly after embryo transfer into recipients (Wang and Dey, 2006), this strongly suggests that the embryos were reabsorbed. Although switching to Rag1$^{-/-}$ blastocysts and to blastocyst injections produced live-born animals (although no chimeras), interpretation of this data is compounded by the fact that the Runx1$^{-/-}$ and Rag1$^{-/-}$ mouse colonies were housed at different sites and, while every effort was made to maintain consistency between the injection groups, it was not possible to control for every potential

variable in the reagents and equipment used. This means that there are many factors which varied between these sets of injections and one or more of them could have contributed to the lack of live-born animals from the Runx1$^{-/-}$ and WT mouse embryos.

The fact that no chimeras were produced from multiple rounds of injection of Rag1$^{-/-}$ embryos using proven stem cell lines indicates a potential problem with the donor cells, rather than with the PCR assay, where the positive controls consistently worked. The DAK31 rat ESC line used for the majority of these injections is well characterised and can produce chimera rates of ~35% when injected into rat embryos (Linda Sutherland, personal communication). However, the data from Kobayashi et al. (2010) indicate that interspecies chimera generation is significantly less efficient than intraspecies chimera production and that chimera rates can vary greatly between lines. Interestingly, when rat ESCs were injected into mouse blastocysts, only ~10% of the animals produced were chimeric as opposed to the ~35% chimera rate seen when a rat iPSC line was used in mouse embryos. It is possible that the combination of the DAK31 cells and the Rag1-null embryos results in a poor chimera rate and that, by simply carrying out more injections or trying a range of different rat ES or iPS cell lines, a viable rat-mouse chimera would have been obtained.

It would also be of use to see how much the donor cells contribute to embryonic development after the blastocyst stage. Although analysis of embryos injected with GFP ESCs indicated that the donor cells were colocalising with the ICM of the blastocysts, it would be informative to culture the embryos to the egg cylinder stage as described by Bedzhov et al. (2014) to ensure that the cells are still retained in the embryo and interacting with the host cells and not ejected from the embryo.

3.4.2. Using partially differentiated cells as donor cells

Another potential alternative for a donor cell source would be to use HSCs or partially differentiated pluripotent stem cells. Indeed, partially differentiated mouse ESCs expressing the markers AA4.1+ and B220- have been shown to be capable of reconstituting the immune system of Rag1$^{-/-}$ mice (Potocnik et al., 1997), demonstrating that the vacant lymphoid niche in these animals can be colonised in intraspecific chimeras. However, there are currently no robust protocols to induce pluripotent stem cells to differentiate into haematopoietic precursors and most successful protocols make use of transgenes to influence cellular differentiation (Tashiro

et al., 2012), something which would ideally be avoided if this method is ever to be used to produce clinical products.

Primary HSCs therefore present a more attractive source of donor cells, at least for the establishment of the technique in rodents and large animals, and have already been used to establish interspecific chimeras. Transplantation of human HSCs into mid-gestation pig embryos resulted in animals which showed a low level of chimerism in the bone marrow 18 months after birth (Ogle et al., 2009), while injection of human HSCs into the liver of sheep foetuses also produced low-level chimeras (Abe et al., 2014). The fact that these chimeras can be produced is a promising indicator that future blastocyst complementation experiments may be a success. However, it should be noted that chimerism in the sheep, which persisted up to 40 months after birth, was only observed when donor cells had undergone forced transient expression of HoxB4 prior to injection (Abe et al., 2014) which, as discussed, is not desirable in for the purposes of producing blood for clinical patients.

Another issue with these chimeras is that any chimerism observed ultimately disappeared from the animals. It is unclear why the donor human HSCs were eventually lost from the bone marrow, but it is potentially due to incompatibilities between the donor cells and the recipient haematopoietic niche, which is either incapable of permanently sustaining the human HSCs or allows the recipient HSCs to outcompete the donor HSCs. An alternative view could be that these donor HSCs could also persist for as long as they did because of supporting factors provided by the host HSCs, a cell population which would not be present in Runx1$^{-/-}$ animals. The compatibility between donor cells and a vacant recipient niche needs to be understood in greater detail.

3.4.3. Interspecies compatibility in the haematopoietic niche

The stem cell niche is the physical environment surrounding a stem cell population which regulates the characteristics of the inhabiting stem cells (Schofield, 1978). HSCs are dependent on the surrounding niche for their development and even require the support from stromal cells for growth *in vitro* (Dexter et al., 1977). It is important to realise that HSCs travel through a number of niches during development, from the dorsal aorta to the foetal liver and then on to the bone marrow. In addition, the bone marrow niche itself contains microenvironments which maintain different states of the HSCs such as quiescence and proliferation. Therefore,

a donor and recipient species used for blastocyst complementation require compatibility in all the niche microenvironments. The many different components of the bone marrow niche, from the different cell types to cytokines and physical factors, have been reviewed extensively (Wang and Wagers, 2011, Sarrazin and Sieweke, 2011, Mendelson and Frenette, 2014).

The potential for incompatibility between donor HSCs and a recipient niche was investigated experimentally by Butler et al. (2010), where murine HSCs were cultured on human endothelial cells which provide numerous factors known to be necessary for HSC maintenance. Only one factor, sKitL, was found to not be cross-compatible between the two species, suggesting that there exists a high degree of conservation in HSC niches between species but that some factors may block niche colonisation. The current rise of systems biology may also help to find possible incompatibilities between donor cells and recipient niches in that signalling pathways and transcriptional networks can be compared across multiple niches in the same species (i.e. the dorsal aorta, the foetal liver and the bone marrow) (Charbord et al., 2014) and across different species (Boyle et al., 2014, Gerstein et al., 2014).

One solution to this issue of interspecies compatibility would be to produce animals with humanised stem cell niches. The feasibility of this has been investigated in mice, where artificial scaffolds and human mesenchymal stem cells were used to create a humanised HSC niche (Groen et al., 2012) in adult mice. Although this could potentially circumvent any interspecies incompatibility in the niche, this process requires immunodeficient recipients and also implants the humanised niche as an addition to the host niche, rather than as a substitution. This results in the support of human HSCs, making this an interesting technique for HSC expansion, but the circulating blood cells are mixed with those of the mouse, meaning that whole blood cannot be produced using these animals.

3.4.4. Restricting donor cell contribution to specific lineages

This work could ultimately lead to the creation of interspecific chimeras using human cells, meaning that important ethical considerations need to be taken into account, as recently reviewed by Shaw et al. (2014). One of the main issues with this area of research is that recipient animals injected with human cells may be inadvertently endowed with 'human features' with the brain, gametes and physical features such as the skin and eyes being of particular concern. Indeed, Han et al. (2013) demonstrated a significant increase in learning ability in mice en-

grafted with human neural precursors. These concerns require both the donor cells and the recipient niche to be carefully chosen. In the case of the Runx1$^{-/-}$ mouse, the fact that Runx1 expression and function is not confined to the haematopoietic niche (as discussed in Section 3.1.3.3) is a potential cause for concern as any injected pluripotent stem cells may rescue deficiencies caused by loss of Runx1 in tissues other than the blood, such as bone and the nervous system in addition to their random contribution to tissues throughout the embryo.

This could potentially be solved by using donor cells with a more restricted potency, such as HSCs. However, HSCs have been shown to be capable of contributing to tissues derived from all three germ layers in the developing embryo (Pessac et al., 2012), including liver tissue (Lagasse et al., 2000) and skeletal muscle (Ferrari et al., 1998, Gussoni et al., 1999). HSCs also express cell markers associated with 'stemness', including Oct4, Sox2 and Klf4; three of the four factors used to make iPS cells (Takahashi and Yamanaka, 2006), as well as genes associated with early progenitor cells in each of the three germ layers (Pessac et al., 2011). This means that injection of HSCs into a recipient embryo does not guarantee that donor contribution will be limited to the haematopoietic niche as they do not seem to preferentially colonise their 'intended' niche (Pessac et al., 2012).

An alternative method for limiting donor cell contribution to chimeras has been recently published by Kobayashi et al. (2014), using the Pdx1-null animals first described in Kobayashi et al. (2010) (detailed in Section 3.1.2.3). An inducible Mixl1 transgene was inserted into donor pluripotent stem cells and Mixl1 transiently expressed both prior to and after injection of the donor cells into a recipient embryo. Mixl1 expression pushes cells to preferentially contribute to tissues and organs derived from the endoderm rather than any ectodermal or mesodermal lineages. However, in addition to the pancreas fully consisting of donor-derived cells, as intended, the researchers also found increased contribution of the donor cells to other organs, such as the heart. This, coupled with the use of a transgene, demonstrate that there is some way to go before large animal interspecific chimeras will be able to produce organs and tissues for transplant into humans.

4. Using genome editors to produce myostatin-knockout sheep

4.1. Introduction

4.1.1. Skeletal muscle development and characteristics

Skeletal muscle is derived from specialised stem cells, known as satellite cells. Satellite cells are located within the skeletal muscle at the boundary between a mature myotube and the extracellular matrix (Mauro, 1961) and are generally found in a quiescent state, similar to other tissue-specific stem cells. Satellite cells exit quiescence when there is an active need to produce more differentiated muscle cells (recently reviewed in detail by Wang et al. (2014d)). Upon activation, satellite cells differentiate into myoblasts, which subsequently fuse together to form myotubes. These myotubes are then arranged into bundles called myofibres, which form the basic unit of a muscle (Gilbert, 2010). This process is shown in Figure 4.1.

Myofibres have been historically classified as fast twitch or slow twitch, depending on their speed of contraction (Peter et al., 1972). Fast twitch fibres predominantly express MHC Type II fibres and generally have a larger diameter than slow twitch fibres, which express more MHC Type I fibres (Pette and Staron, 2000). Fast twitch fibres are more likely to produce ATP by glycolytic means while slow twitch myofibres, which contain a higher number of mitochondria than fast twitch fibres, produce ATP by oxidative means. Producing ATP by glycolysis allows fast twitch fibres to produce ATP more quickly than slow twitch fibres, but at the cost of increased lactic acid production, meaning that fast twitch fibres fatigue more rapidly than slow twitch (Levy et al., 2006, Hall and Guyton, 2011).

Muscle development and the balance of fast to slow twitch fibres in myofibres is regulated by myostatin, amongst other factors.

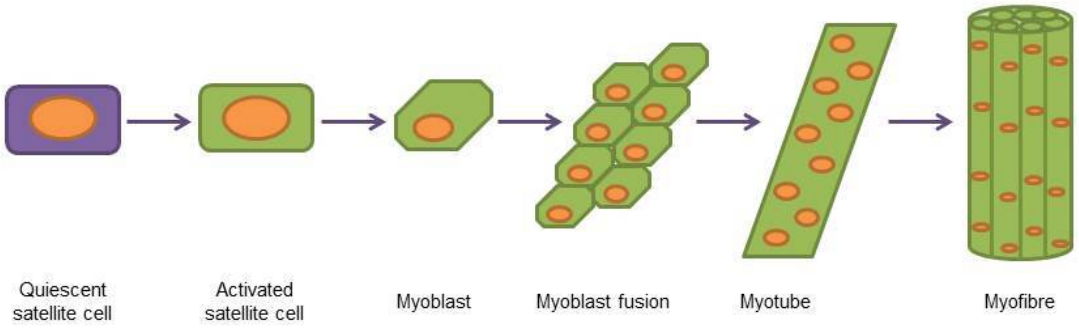


Figure 4.1.: Muscle development from satellite cell to myofibre

4.1.2. Myostatin

The role of myostatin as a negative regulator of muscle growth was first identified by the striking phenotype of the myostatin-null mouse (McPherron et al., 1997), described in detail in Section 4.1.3. Myostatin is a member of the TGF-β family (McPherron et al., 1997) and is most closely related to BMP11 (McPherron et al., 1999), which also regulates growth and development in a number of tissues and is expressed in a wider range of tissues than myostatin (Lee and Lee, 2013). Myostatin is highly conserved across species (McPherron et al., 1997, McPherron and Lee, 1997), which is indicative of its important regulatory role.

Myostatin expression begins during embryogenesis in the myotome compartment of developing somites from stage E9.5. Expression continues into adulthood, where it is primarily expressed in skeletal muscle (McPherron et al., 1997), although there is also limited expression in adipose tissue, where myostatin negatively regulates brown adipocyte differentiation (McPherron et al., 1997, McPherron and Lee, 2002, Braga et al., 2013).

The myostatin gene contains three exons, which, once transcribed and translated, form an peptide of 376 amino acids in length. Transcription of myostatin is initiated by the binding of FoxO1 and Smad transcription factors to the promoter region (Allen and Unterman, 2007). The full-length peptide is processed into three parts; amino acids 1-23 form the signal peptide for secretion from skeletal muscle into the circulatory system, amino acids 24-266 form the propeptide and amino acids 267-376 form the mature protein (McPherron et al., 1997, Lee and McPherron, 2001). Like other TGF-β family members, the full-length peptide is processed firstly to remove the signal peptide, then the propeptide is cleaved from the mature protein by

BMP-1/tolloid metalloproteinases (Wolfman et al., 2003), which allows activation of the mature myostatin protein. During processing, which occurs in skeletal muscle, mature myostatin proteins dimerise via disulphide links (McPherron et al., 1997). The propeptide appears to play a role in ensuring that the mature protein is folded correctly and regulates protein expression by holding myostatin in a latent, inactive form (Lee and McPherron, 2001, Hill et al., 2002). Myostatin can be also be inhibited by molecules such as follistatin and follistatin-like proteins like FLRG (Hill et al., 2002, Cash et al., 2009, 2012), GASP-1 and GASP-2 (Lee and Lee, 2013). These inhibitors are thought to prevent myostatin from binding to its receptors (Lee and McPherron, 2001).

Latent and activated myostatin protein circulates in the blood (Zimmers et al., 2002) and, once activated, acts on cells via autocrine signalling (McPherron et al., 1997, Rios et al., 2004). Myostatin-mediated signalling can occur via the TGF-β or MAPK signalling pathways. TGF-β signalling is initiated by the binding of myostatin to the activin type II receptor ActRIIB (Rios et al., 2004) and the subsequent binding and activation of the type I receptors ALK4 and ALK5. ALK4 and ALK5 activate the signal transducers Smad2 and Smad3 by phosphorylation (Rebbapragada et al., 2003). This ability of myostatin to activate Smad proteins suggests that it may have a role in stimulating its own transcription. Myostatin also signals through the MAPK pathway via p38 and the TAK1-MKK6 cascade (Philip et al., 2005). Myostatin-induced MAPK signalling is independent from TGF-β signalling.

Myostatin's regulation of muscle growth occurs through two main mechanisms. Firstly, it induces satellite cell quiescence via the upregulation of the Cdk inhibitor p21 and decreasing Cdk2 protein levels, which prevents cells from entering S phase (McCroskery et al., 2003). It also reduces myoblast proliferation, causing myoblast cells to remain in G1 phase and reduces the expression of MyoD and myogenin proteins, which are two key drivers of muscle development. (Joulia et al., 2003).

In addition to its role in controlling muscle growth, myostatin also regulates the type of myofibres which develop. Although it is expressed throughout skeletal muscle, it is differentially expressed in myofibres in fast and slow muscles, where it inhibits the development of the MHC Type II myofibres found in fast-twitch muscles and encourages the expression of slow-twitch MHC Type I myofibres. Myostatin also regulates muscle growth by inducing satellite cell quiescence via the upregulation of the Cdk inhibitor p21 and decreasing Cdk2 protein levels,

preventing cells from entering S phase (McCroskery et al., 2003).

4.1.3. Myostatin-knockout phenotype

The striking phenotype of the myostatin-knockout mouse was first described by McPherron et al. (1997). Knockout animals were created by replacing exon 3 of the myostatin gene, which encodes the mature protein, with a neomycin cassette. The resulting knockout animals had a total body size which was about 30% bigger than that of wildtype animals and individual muscles which weighed 2-3 times more than wildtype counterparts (McPherron et al., 1997), leading to the naming of the phenotype as double muscled. Sections from muscles of myostatin-null animals showed than this increase in muscle was due to a combination of hypertrophy (knockout muscles had a total cell count which was 86% higher than wildtype muscles) and hyperplasia (myostatin-knockout muscle fibres had a larger diameter than wildtype controls) (McPherron et al., 1997). Double muscled animals have increased absolute muscle power but reduced endurance (Baligand et al., 2010, Matsakas et al., 2010) and exhibit a 70% reduction in adipose tissue accumulation despite having similar food intake and metabolic rates to wildtype animals (McPherron and Lee, 2002).

A second double muscled mouse line has also been characterised. The Compact mouse arose from a breeding program selecting for high carcass protein content (Varga et al., 1997) and displays a similar hypermuscular phenotype to the myostatin-null mouse described above. The Compact phenotype is cause by a 12bp deletion which causes a deletion of four amino acids and the substitution of a further amino acid in the propeptide. Given the central role of the propeptide is regulating the mature myostatin protein, it is likely that the double muscling of the Compact mouse is due to the mutated propeptide either failing to prevent misfolding of the mature protein or sequestering an increased amount of myostatin protein (Szabo et al., 1998). Given the nature of the mutation, it is likely that the Compact mouse represents the effects of a substantial knockdown of the myostatin protein, rather than a complete knockout.

Dominant negative myostatin mutants, where the myostatin signalling pathway has been experimentally disrupted, also exist. Examples include ActRIIB mutants (Lee and McPherron, 2001) and mutants where the cleavage site between the propeptide and mature protein has been disrupted (Zhu et al., 2000). However, some of these mutants differ from the myostatin-null mouse in that they show a phenotype due to hypertrophy only with no contribution from

hyperplasia (Zhu et al., 2000).

Several mechanisms which cause the double muscle phenotype have been elucidated. The hypertrophy of the muscles is at least partially due to an increase in satellite cell proliferation, leading to production of more myoblasts. The proliferation rate of myoblasts themselves is also increased by the loss of myostatin (McCroskery et al., 2003). An increased number of myoblasts also contributes to myofibre hyperplasia, as an increased number of myoblasts fuse together to create a single myofibre. Hyperplasia is also partly driven by an increase in protein synthesis via upregulation of Akt/mTOR signalling (Rodriguez et al., 2011, Trendelenburg et al., 2009).

The increased power and fatigability of double muscled animals stems from myostatin's regulation of the types of MHC molecules expressed in myofibres. Muscle fibres from knockout mice are more likely to differentiate into those containing MHC IIB fibres, which confer fast-twitch properties (Amthor et al., 2007, Girgenrath et al., 2005). This increases muscle power, but at the expense of fibres containing MHC Type I, or slow-twitch, fibres, which increase endurance. This switch from MHC Type I to MHC Type II fibres is influenced by downregulation of MEF2C and upregulation of MyoD caused by loss of myostatin (Hennebry et al., 2009). Loss of myostatin also causes more myofibres to take on glycolytic characteristics and lose oxidative properties, including a reduction in mitochondrial numbers and respiration rate (Amthor et al., 2007, Baan et al., 2013, Mouisel et al., 2014). Producing energy through glycolytic rather than oxidative means has a greatly reduced efficiency, thereby increasing the ATP cost of contraction (i.e. more ATP is required to generate the same amount of force compared to a WT myofibre) (Giannesini et al., 2013). However, oxidative activity can be increased following a regular exercise regimen (Matsakas et al., 2010).

4.1.3.1. The myostatin-knockout phenotype in livestock animals

Based on this description of the double muscled phenotype caused by a loss of myostatin in mice, it is clear to see why a similar mutation in livestock animals such as cattle and sheep would have a high agricultural value. Double muscled animals yield more and leaner meat and, indeed, there are already examples where naturally occurring inactivating mutations in myostatin have been selected for in certain breeds. The best known example of this is the Belgian Blue cattle breed, which shows a 20-25% increase in muscle size over other cattle

breeds (McPherron and Lee, 1997) . The double muscle phenotype of the Belgian Blue is caused by an 11bp deletion within the third exon of myostatin, which causes a frameshift and the production of a severely truncated mature protein only twenty amino acids long (McPherron and Lee, 1997, Grobet et al., 1997). This mutation completely abrogates myostatin's function, rendering the animal a true functional knockout.

Double muscling is also seen in Piedmontese cattle and Texel and Norwegian White sheep. Piedmontese cattle have a G to A transition in the third exon that causes a cysteine to tyrosine substitution. This is thought to disrupt the structure of the mature protein, leaving it non-functional as the substituted cysteine is highly conserved, not only across myostatin proteins from different species, but across all TGF-β family members, suggesting the importance of the cysteine in the protein (McPherron and Lee, 1997).

Like the Piedmontese, the Texel mutation is a G to A transition in the 3' UTR of the myostatin gene, creating an illegitimate target site for the microRNAs *mir1* and *mir206* (Clop et al., 2006, Takeda et al., 2010). This results in a RNAi-induced reduction in the levels of myostatin mRNA to around one-third of the levels seen in wildtype animals via RNAi (Clop et al., 2006). This increases muscle-related traits and decreases fat-related traits in Texel animals but, unlike knockout animals, the mutation has no significant effects on total body weight or growth traits (Johnson et al., 2009, Masri et al., 2011). Norwegian White sheep have the same G to A transition in the myostatin 3'UTR and some animals also have a 1bp deletion within exon 3. This deletion produces a truncated mature protein however, as the extent of truncation is significantly less than that seen in the Belgian Blue, it is not clear whether the protein retains any biological activity (Boman et al., 2009). However, lambs which are homozygous for this deletion tend to die shortly after birth and the mutation is being actively bred out of the population (Boman et al., 2011).

Despite the presence of myostatin knockout mutations in the natural populations of some livestock breeds, it would be of great use to be able to quickly introduce established or new myostatin alleles in other breeds. Current double-muscled animals, such as the Texel and Belgian Blue, are not feasible breeds to be used in many places where livestock farming occurs, for a variety of reasons. Inappropriate environmental conditions, such as an increased outdoor temperature, can limit some breeds' ability to thrive, as they cannot adequately regulate their body temperature. In addition to this, some areas, such as parts of Sub-Saharan Africa, can have

severely restricted access to veterinarians and other animal husbandry resources. This would be a particular problem considering the reproductive issues associated with Belgian Blue cattle, meaning that the productivity benefits of the breed would be significantly negated by the loss of animals during labour. Therefore, the ability to introduce myostatin-knockout mutations into breeds which are well adapted to their local conditions could be of great benefit to local farmers.

Recent publications have described attempts to produce sheep and goats using various tools such as RNAi or genome editors such as ZFNs or CRISPRs (Zhang et al., 2013, Zhong et al., 2014, Ni et al., 2014), emphasising the value of the double muscled phenotype. This results chapter presents the use of TALENs and CRISPRs to produce edited myostatin alleles in sheep via embryo microinjection.

4.2. Additional materials and methods

4.2.1. Muscle staining

Four different leg muscles were sampled; the soleus, the tibialis anterior, the gastrocnemius and the peroneus longus (see Figure 4.2). A section of muscle was cut away from the central portion of each muscle and each sample divided for either cryosectioning or wax embedding, as described below. Images of all stained sections were taken at 200x magnification. Three images were taken for each muscle from each animal and the cross-sectional area of each fibre in the images was measured using ImageJ.

4.2.1.1. Muscle cryosections and enzyme staining

Muscle samples to be cryosectioned were embedded in OCT compound (Tissue-Tek) on cork discs before being snap frozen using Cryospray (VWR international) and placed into liquid nitrogen. Samples were then recovered from the liquid nitrogen for long-term storage at -80ºC. 10μm serial sections were cut in a cryostat machine and placed onto poly-L-lysine slides. Freshly cut sections were allowed to dry at room temperature for one hour prior to staining.

Figure 4.2.: Image showing the location of the gastrocnemius (G), the peroneus longus (P), the soleus (S) and the tibialis anterior (TA) within the leg.

4.2.1.2. ATPase staining

The following solutions were prepared in advance of ATPase staining;

Formal-cacodylate sucrose fixative 20g of paraformaldehyde was added to 100ml of distilled water and dissolved by heating to 60ºC using a magnetic stirrer. Drops of 10% sodium hydroxide were added until the solution cleared then the following reagents were added; 62.5g of sucrose, 10.7g of sodium cacodylate, 0.625g of sodium pyrophosphate and 400ml of distilled water. The pH of the solution was then adjusted to 7.4 and the fixative stored at 4ºC.

1% ammonium sulphide Just before use, 3ml of 20% ammonium sulphide (Fisher Scientific) was added to 60ml of distilled water.

Glycerin jelly 30g of glycerol jelly was warmed in 50ml of distilled water and 20ml of glycerol until the jelly had fully dissolved. The jelly was then aliquoted and stored at 4ºC. Individual aliquots were warmed to 37ºC before use.

ATPase incubation medium 1.21g of Tris base (Fisher Scientific) and 0.2g of calcium chloride were added to 100ml of distilled water and the pH adjusted to 9.5. Just before use, 40ml of the solution was aliquoted and 60mg of ATP (disodium salt) was added to the aliquot.

After drying, fresh cryosections were fixed in formal-cacodylate sucrose fixative at 4ºC for 60-90 minutes. Slides were then washed once in distilled water before being placed in a glass Coplin jar containing ATPase incubation medium and incubated in a water bath set to 37ºC for 45-60 minutes.

After incubation, the slides were washed twice in a 1% calcium chloride solution (each wash was three minutes long), then transferred to a 2% cobalt nitrate solution for three minutes. Slides were then washed twice in distilled water and briefly plunged into 1% ammonium sulphide to develop. Excess ammonium sulphide was then removed by two washes in distilled water and sections mounted using glycerin jelly and coverslips.

4.2.1.3. NADH staining

The following solutions were prepared in advance of NADH staining;

0.1M phosphate buffer 8.5g of sodium chloride, 1.07g of disodium phosphate and 0.39g of monosodium phosphate were dissolved in 1 litre of distilled water.

NADH incubation medium A 10mg nitrotetrazolium blue tablet was added to 10ml of 0.1M phosphate buffer (described above) and vortexed until the tablet was fully dissolved. The solution was then stored until needed, at which point 8mg of NADH was added to the solution just before use.

After drying, slides with fresh cryosections were covered in NADH incubation medium and incubated at 37ºC for 45 minutes. Slides were then washed twice in distilled water and mounted using glycerin jelly before being covered with a coverslip and left to dry overnight.

4.2.1.4. Wax embedding and HE staining

Muscle samples for wax embedding were fixed in 4% paraformaldehyde overnight. Fixation was carried out on a rocking platform at 4ºC. The following day, the samples were washed using the following protocol and stored at 4ºC in the final 70% ethanol wash.

Wash reagent	Wash time	Wash temperature
PBS	30 minutes	4ºC
30% ethanol	15 minutes	Room temperature
70% ethanol	30 minutes	Room temperature
70% ethanol	30 minutes	Room temperature

Samples were then placed into labelled histology cages and processed for wax embedding using the following protocol.

Reagent	Incubation time
70% ethanol	2 hours
90% ethanol	2 hours
100% ethanol	2 hours
100% ethanol	2 hours
Xylene	2 hours
Xylene	2 hours
Xylene	2 hours
Wax	2 hours
Wax	10 hours
Wax	2 hours

Samples were subsequently embedded in wax blocks and stored at 4ºC. 10μm sections were cut on a microtome and placed onto poly-L-lysine slides. Sections were incubated at 50ºC overnight to melt away excess wax before staining.

H&E staining was carried out using a Leica XL autostainer using the following protocol. After staining, slides were then mounted using Histomount and covered with coverslips.

Reagent	Time
Xylene	5 minutes
Xylene	5 minutes
Xylene	5 minutes
100% ethanol	3 minutes
100% ethanol	2 minutes
95% ethanol	2 minutes
Distilled water	5 minutes
Haematoxylin	3 minutes
Distilled water	3 minutes
Distilled water	3 minutes
Tap water	2 minutes
Distilled water	2 minutes
Eosin	2 minutes
Distilled water	45 seconds
70% ethanol	30 seconds
95% ethanol	30 seconds
95% ethanol	30 seconds
100% ethanol	1 minute
100% ethanol	1 minute
100% ethanol/xylene (50:50 mix)	1 minute
Xylene	1 minute
Xylene	1 minute
Xylene	1 minute

4.3. Results

4.3.1. TALEN and CRISPR design and validation

TALENs (designed by Dr Daniel Carlson) and CRISPRs were designed to target exon 3 of the MSTN gene, which corresponds to the mature protein (see Figure 4.3), and constructed as detailed previously. MSTN TALENs were originally designed against the bovine MSTN locus only and were provided as capped mRNA by Recombinetics. This TALEN mRNA was transfected into OEFs and BEFs using the Neon system and a CelI assay carried out. Evidence of editing was seen in both cell types, with TALEN activity being higher in the BEFs than the OEFs (Figure 4.4), as indicated by the greater intensity of the cleavage products. The amount of exposure in the image means that the amount of editing seen cannot be quantified. An image at a lower exposure would not have clearly shown the activity of the TALENs in the ovine cells. The success of this validation meant that the TALEN mRNA could be used further in embryo injections, detailed in Section 4.3.2.

The MSTN TALENs were subsequently constructed into the ELD/KKR vectors, which allowed TALENs which were specific to either the bovine or ovine MSTN target site to be produced (performed by Mrs Claire Neil). The completed TALENs were then transfected into BEFs or OEFs as plasmid DNA along with a GFP plasmid to assay transfection efficiency (Figure 4.5 (a)). However, unlike the positive results found upon transfection of the MSTN TALEN mRNA, transfection of the plasmids did not lead to the production of any cleavage products in a CelI assay (Figure 4.5 (b)).

To eliminate the possibility that the TALENs may have been inadvertently cloned into the incorrect version of the heterodimeric ELD/KKR vectors so that, for example, both TALENs had been cloned into an ELD vector, the region of FokI containing the ELD and KKR point mutations was sequenced for each TALEN along with the empty ELD and KKR plasmids (Figure 4.6). Aligning the resulting sequences showed that the left and right TALENs were in their intended destination plasmids, with the left TALENs in the ELD vector and the right TALEN in the KKR vector.

The MSTN target sites in the bovine and ovine fibroblasts were sequenced to check for possible mismatches between the target DNA and the TALEN RVDs, although this was thought unlikely as the TALENs had been previously shown to be active in both cell types. The target

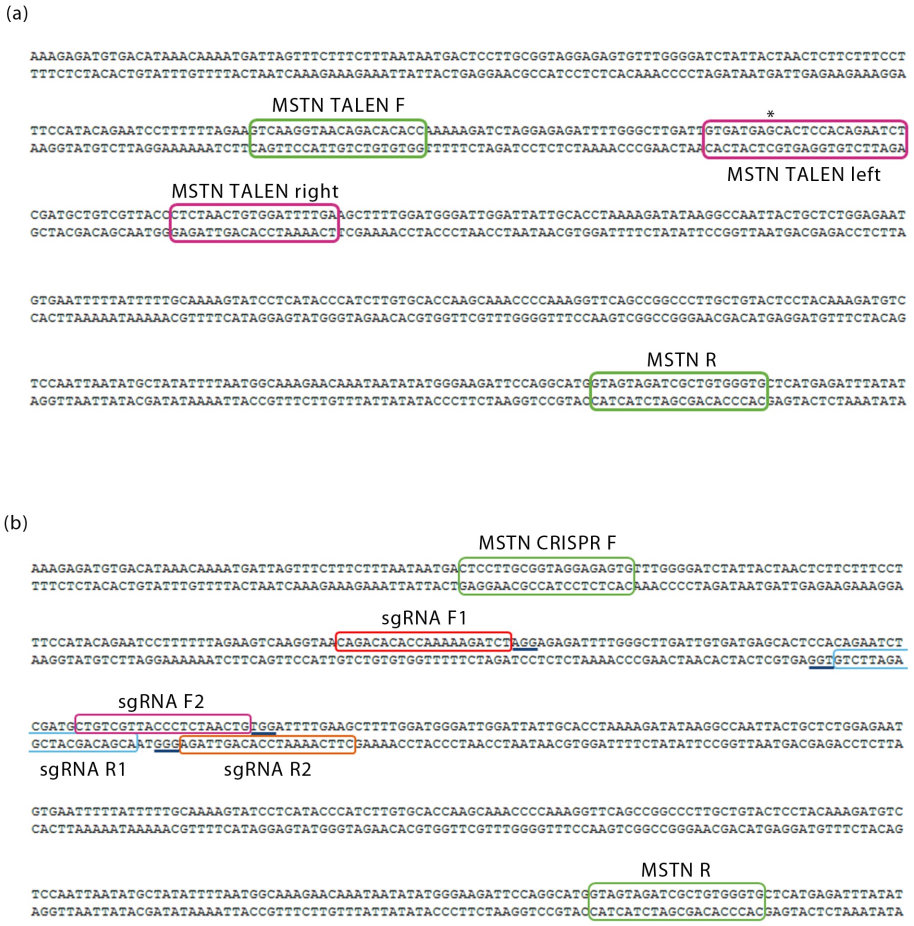


Figure 4.3.: Positions of (a) TALEN and (b) sgRNA binding sites in exon 3 of the ovine myostatin locus. The bovine SNP is highlighted in (a) by an asterisk. PAM sequences for each of the sgRNAs are underlined in blue.

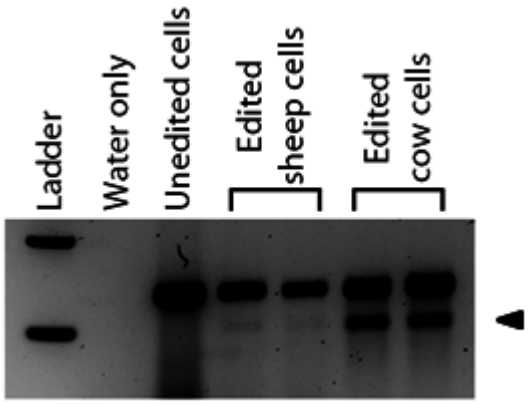


Figure 4.4.: Gel image showing the results of a CelI assay on OEFs and BEFs transfected with MSTN TALEN RNA provided by Recombinetics. Cleavage products are indicated by the arrowhead.

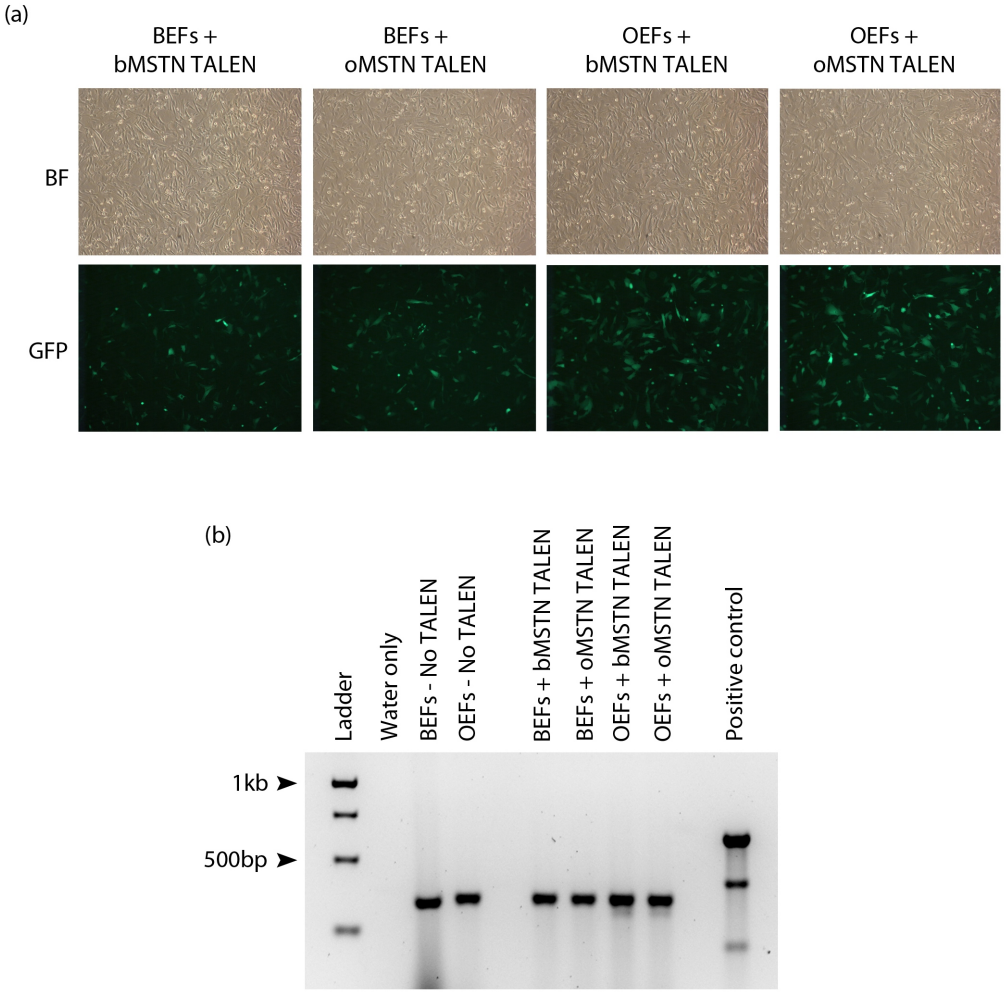


Figure 4.5.: Validation of bovine and ovine MSTN TALENs in the Sangamo ELD/KKR vectors. TALENs targeted to either the bovine MSTN (bMSTN TALEN) or ovine MSTN (oMSTN TALEN) loci were transfected into BEFs and OEFs along with a GFP plasmid to show transfection efficiency. (a) Brightfield and GFP images of transfected BEFs and OEFs 3 days after transfection. Cells were transfected with both versions of the MSTN TALENs to determine whether the difference in efficiency observed in Figure 4.4 was maintained with the new expression vectors. Images at 100x magnification. (b) Results of CelI assay on transfected cells. No evidence of TALEN activity was seen in any of the transfected samples.

sites in both the BEFs and OEFs aligned completely with their respective genomic sequence (Figure 4.7). This, combined with the results of other troubleshooting steps taken for the Rag1 TALENs (see Figure 3.15), left no clear reasons why the TALENs had suddenly lost their activity.

The MSTN sgRNAs were constructed in the pSpCas9(BB) vector and, as with the MSTN TALENs, tested by transfection into BEFs and OEFs (Figure 4.8). Although the expression of

Figure 4.6.: Sequence alignments of the FokI region of the myostatin TALENs which contains the heterodimer ELD/KKR mutations. The ELD/KKR point mutations are highlighted in green boxes

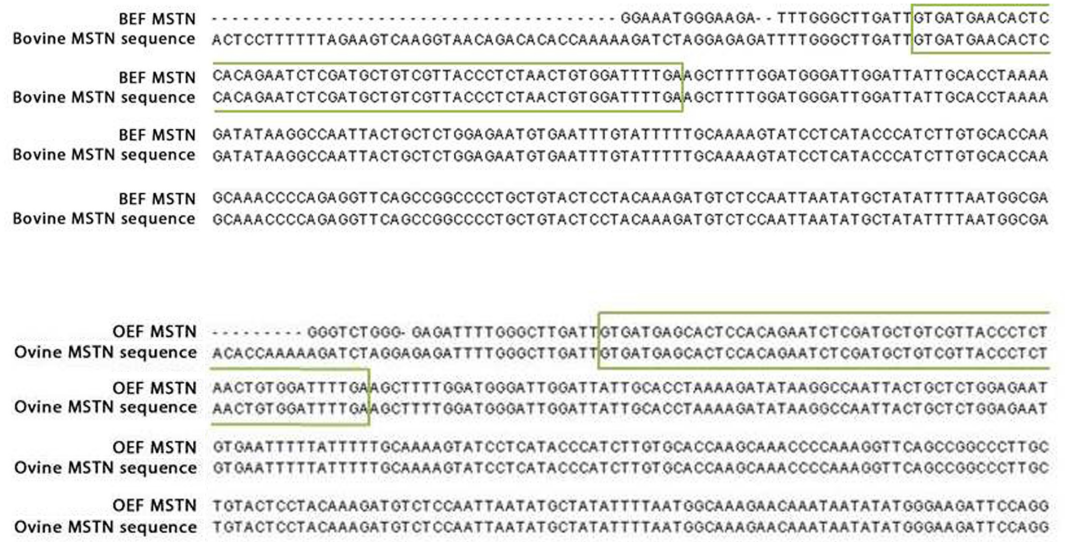


Figure 4.7.: Sequence alignments of myostatin TALEN target sites sequenced from OEF or BEF DNA aligned against their counterparts in the ovine or bovine genome. Areas targeted by the myostatin TALENs are highlighted in green boxes

GFP in the cells showed that the plasmid had been successfully transfected, no editing activity was observed following a CelI assay. To ensure that the transfection protocol was appropriate for use with the CRISPR/Cas9 system, repeat transfections were carried where BEFs and OEFs were co-transfected with the MSTN F1, F2, R1 and R2 CRISPR/Cas9 plasmids alongside a separate transfection where BEFs were transfected with a plasmid containing a sgRNA sequence targeted against Nanos2 (provided by Dr Simon Lillico). This sgRNA had been previously shown to be capable of editing. The results of a T7 assay carried out on these cells

is shown in Figure 4.9. Although co-transfection of the MSTN sgRNAs did not result in any editing activity, cleavage products can be seen in the lane containing DNA isolated from BEFs transfected with the Nanos2 sgRNA, showing that the transfection protocol is suitable for use with CRISPR/Cas9 plasmids.

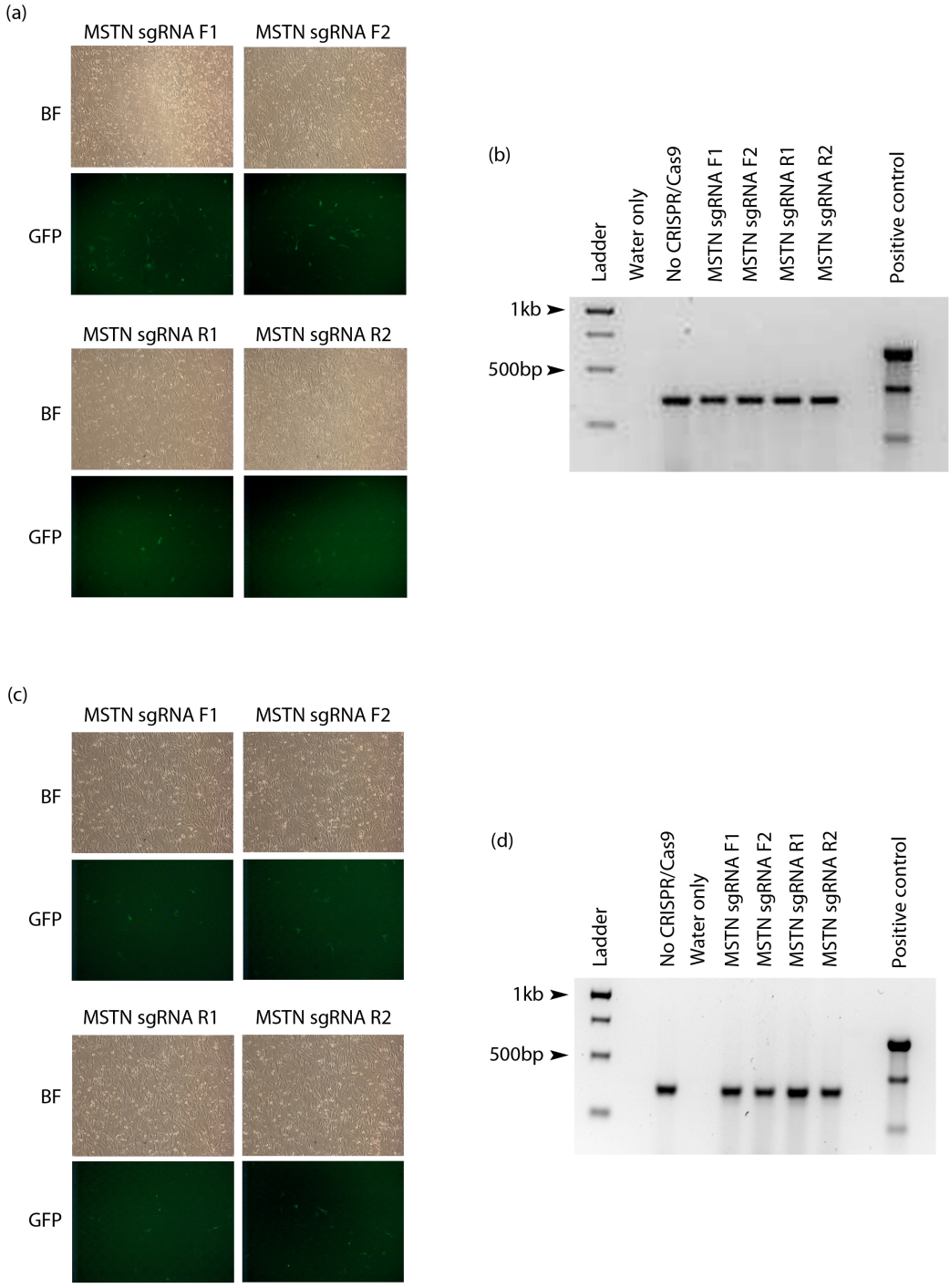


Figure 4.8.: Validation of MSTN sgRNAs in BEFs and OEFs. (a) Brightfield and GFP images of OEFs transfected with sgRNAs cloned into the pSpCas9(BB) vector. Images at 100x magnification. (b) Results of CelI assay on transfected OEFs. (c) Brightfield and GFP images of BEFs transfected with sgRNAs cloned into the pSpCas9(BB) vector. Images at 100x magnification. (d) Results of CelI assay on transfected BEFs. Note that there is no evidence of CRISPR activity in either (b) or (d).

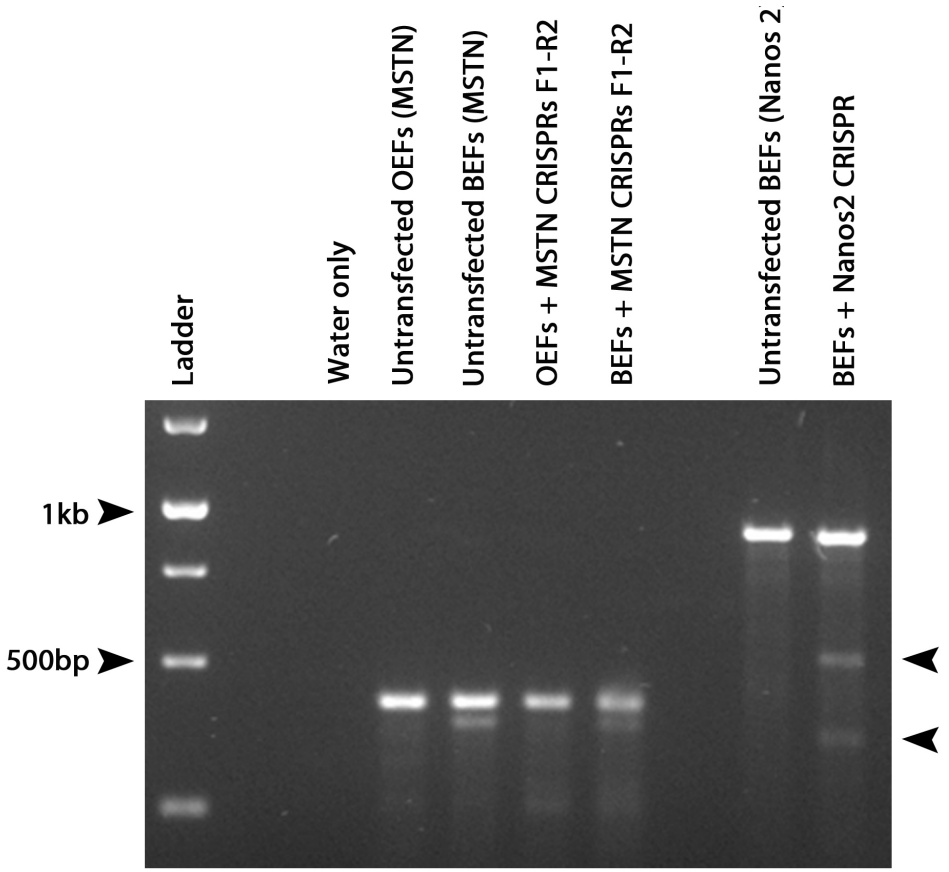


Figure 4.9.: Gel image showing the results of a T7 assay on ovine and bovine fibroblasts transfected with plasmids containing all four MSTN sgRNA sequences (F1, F2, R1 and R2) in addition to bovine fibroblasts transfected with a plasmid containing the Nanos2 sgRNA sequence. Although no evidence of editing was seen in any of the cells transfected with the MSTN CRISPRs, cleavage products indicating editing activity in BEFs transfected with the Nanos2 CRISPR are shown by arrowheads.

4.3.2. Embryo injections

Ovine oocytes were prepared for injection and injected with either MSTN TALEN mRNA or sgRNAs. These injections are summarised in Table 4.1. MSTN TALEN mRNA was injected at a concentration of 2ng/µl. Zygotes injected with MSTN CRISPRs were injected with all four sgRNAs, with each individual sgRNA at a concentration of 10ng/µl, in addition to capped and polyadenylated Cas9 mRNA (PNA) at a concentration of 100ng/µl. All injections had a volume of 2-5pl injected into the cytoplasm.

In addition to the embryos transferred into recipients, some embryos injected with the MSTN sgRNAs were cultured to blastocyst stage and analysed by PCR amplification and subsequent sequencing of the MSTN locus for the presence of editing events. No edited alleles were

detected in any of the 28 embryos analysed (data not shown).

DNA from lambs resulting from these injections was sequenced to determine whether any of the animals were edited. Out of 16 lambs born, two lambs, one injected with MSTN TALENs and one injected with MSTN sgRNAs, were found to have edited myostatin alleles. The analysis of these animals is presented in Sections 4.3.3 and 4.3.4.

Construct	Embryos injected	Cleaved embryos	Embryos transferred	Pregnancies	Lambs	Edited lambs
MSTN TALEN (Recombinetics RNA)	113	61 (54%)	27 (24%)	8	12 (4 dead)	1
MSTN TALEN (Sangamo ELD/KKR RNA)	56	28 (50%)	5 (9%)	0	-	-
MSTN sgRNAs (F1, F2, R1, R2)	23	8 (34%)	7 (31%)	0	-	-
MSTN sgRNAs (F1, F2, R1, R2) and PPT1 sgRNA	40	30 (75%)	15 (38%)	3	4	1

Table 4.1.: Table summarising microinjections carried out in ovine embryos with genome editors designed to target myostatin. Percentage values for cleavage and blastocyst formation are given in relation to the number of embryos injected.

4.3.3. Identification and analysis of TALEN-edited lamb

The edited lamb produced by the injection of MSTN TALEN mRNA was found to be a mosaic with one edited allele (Figure 4.10 (b)), as the second trace on the chromatogram had a lower signal intensity than the wildtype trace, excluding the possibility that the animal was a heterozygote. It is therefore clear that the editing event took place in a single cell after the zygote had already undergone cell division. Cloning of PCR products confirmed that a 3bp deletion had occurred at the locus although, due to the slightly repetitive nature of the sequence immediately surrounding the cut site, it is unclear exactly which three nucleotides have been deleted (the possible deletion events are shown in Figure 4.10 (d)). However, the result of each potential event is the deletion of a single amino acid, arginine 283, from the myostatin protein.

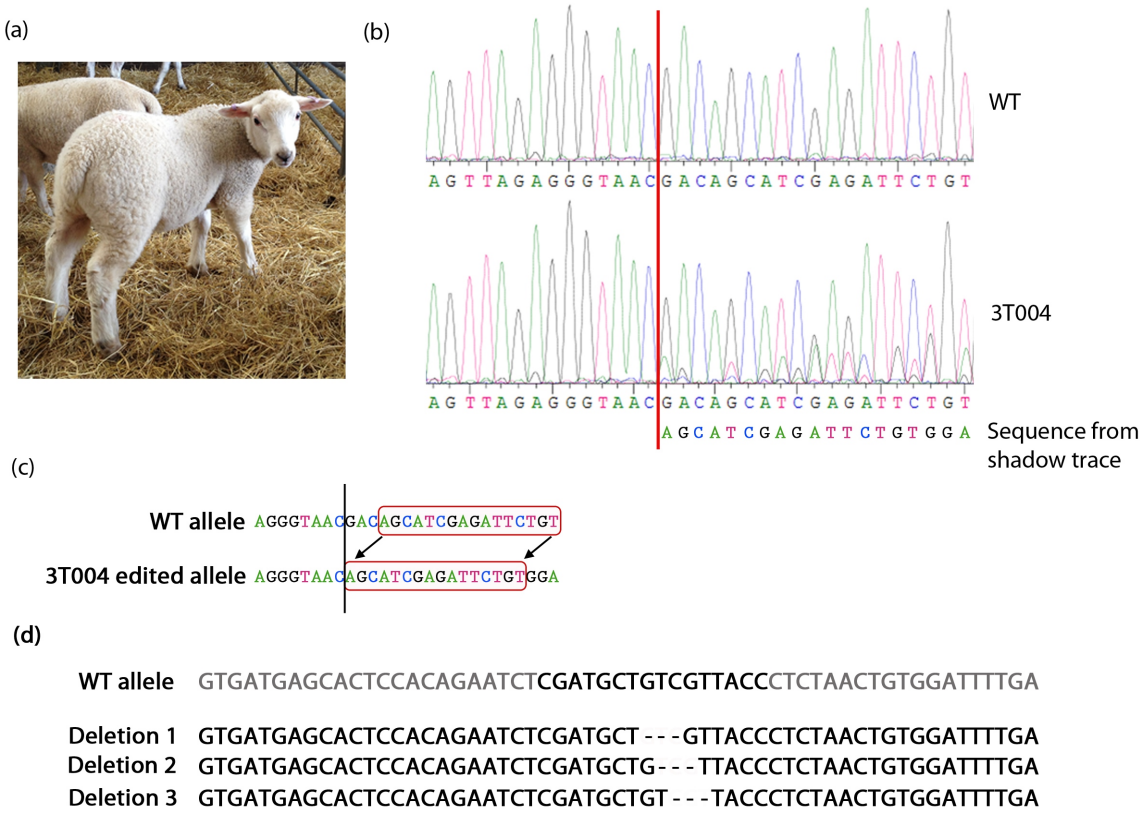


Figure 4.10.: Identification and analysis of myostatin-edited lamb produced using TALENs. (a) Image of genome-edited lamb (ID = 3T004). (b) Comparison of sequence chromatograms from wild-type and edited animals. The red line marks the beginning of the 'shadow' trace, indicating gene editing. The sequence shown by the 'shadow' trace is shown under the chromatograms. The relatively weak signal from the shadow trace compared to the 'wildtype' trace indicates that the animal is a mosaic, rather than a heterozygote. (c) Comparison of wildtype and edited DNA sequences. The black line indicates the point where the shadow trace appears on the edited chromatogram and where the two sequences diverge from each other. The red boxes indicate sequence common to the two alleles and show the 3bp deletion caused by the editing event. (d) The three possible deletions resulting from gene editing. Note that sequences in (b) and (c) are from sequencing reactions using the myostatin reverse primer and are the reverse complement of the sequences shown in (d). Figure (d) adapted from Proudfoot et al. (2014).

The TALEN-edited lamb was not used in a breeding programme due to the fact that the edited myostatin allele was not predicted to be severely affected by the loss of arginine 283, in addition to the animal's mosaic status, which would mean that few offspring, if any, would inherit the edited allele. It was therefore decided to cull the animal and perform a more detailed analysis of the muscle phenotype by measuring the cross-sectional area of muscles fibres. Samples were taken from four calf muscles; the gastrocnemius, the peroneus, the soleus and the tibialis anterior. These muscles were selected to represent the range of different muscle properties found in the leg, in case the edited myostatin protein had differing effects on different muscles as previous work has found that different muscles within the same animal exhibit hypertrophy and hyperplasia to different degrees upon knockout of myostatin (Baan et al., 2013). The tibialis anterior and peroneus predominantly consist of fast-twitch fibres (Baldwin and Tipton, 1972, Ariano et al., 1973, Xu et al., 1998), while the gastrocnemius and soleus contain higher proportions of slow-twitch fibres (around 50% and 70% respectively) (Edgerton et al., 1975).

Muscle samples were either fixed for wax embedding or frozen in OCT for cryosectioning. Originally, H&E staining was used to make the boundaries of each myofibre clear for the quantification of cross-sectional areas. However, wax embedding of the muscle samples caused significant shrinkage of the fibres to occur, which would have impacted on the results (Figure 4.11). Cryosections of muscle samples stained for either ATPase or NADH activity did not appear to shrink to the same extent (Figure 4.11) (although some slight shrinkage was observed in the sections stained for ATPase activity), so were used in place of the H&E sections in the analysis.

Cross-sectional areas of myofibres were calculated using ImageJ. Each muscle has an n of between 52 and 127. The data are presented in Figure 4.12. It is clear that the mean cross-sectional areas of myofibres in the edited animal are amongst the smallest out of the five animals analysed in total and it is therefore safe to anticipate that the edited myostatin allele in this animals does not exert a phenotypic effect. Because only one edited animal was available for analysis, two-tailed t-tests were carried out by comparing each control muscle to the equivalent edited muscle. Although there are significant differences between some of the muscle samples, the fact that there is only one edited animal means that the degree of variation seen between the control animals cannot be adequately accounted for when analysing the single edited animal.

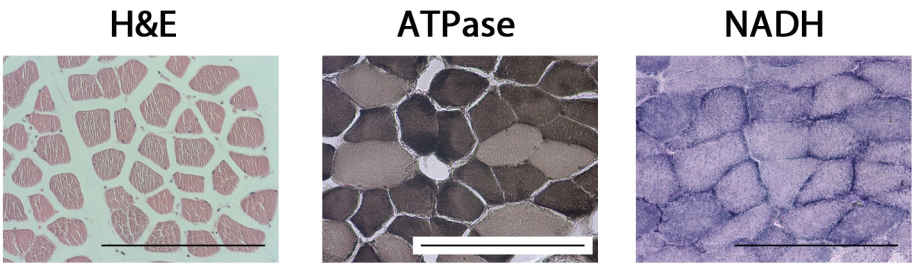


Figure 4.11.: Comparison of the effects of different staining methods on sections of tibialis anterior muscle from a control animal. Scale bars represent 100μm. Darker staining in the ATPase and NADH sections indicate increased ATPase activity and oxidative activity, respectively. Note that the fibres in the H&E stained section (left image) have shrunk significantly in comparison to those in the ATPase and NADH-stained sections (centre and right images), making the H&E sections unsuitable for analysis of myofibre size.

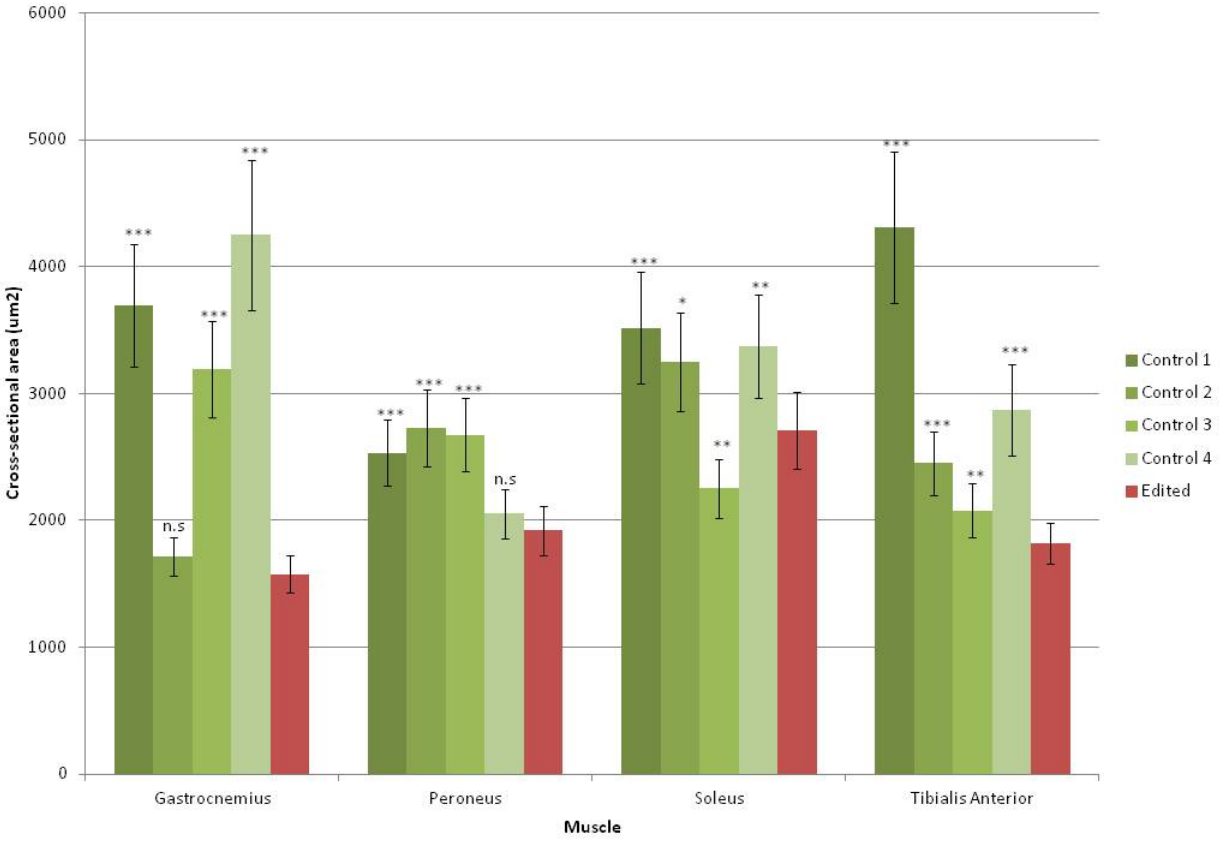


Figure 4.12.: Graph comparing the mean cross-sectional areas of muscle fibres taken from the legs of the TALEN-edited sheep and four unedited controls. T-tests were carried out comparing each control animal to the edited animal in each muscle. * = P <0.05, ** = P <0.01, *** = P <0.001

4.3.4. Identification of CRISPR-edited lamb

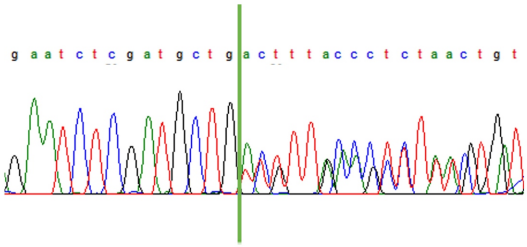
Sequencing of earclip DNA from the lambs resulting from injections of the MSTN CRISPRs showed one animal to be heterozygous for an editing event (Figure 4.13 (b)). Subsequent sequencing of cloned PCR products (Figure 4.13 (c)) showed that the majority of sequenced PCR products harboured a 20bp deletion which is situated 16bp downstream of the deletion event found in Belgian Blue cattle (Figure 4.13 (d)). In addition to this, two sequences also contain single nucleotide changes at sites upstream or downstream of the 20bp deletion (highlighted in Figure 4.13 (c)). Given that sequencing of other PCR products using the DreamTaq polymerase did not show the introduction of SNPs at such a high frequency, it is anticipated that these base changes are the result of editing events which occurred after cell division in the embryo.

The 20bp deletion event found in this edited animal converts a TGT codon, which codes for cysteine, to a TGA stop codon. As a result of this immediate truncation of the protein, this animal is predicted to produce a shorter myostatin protein than that found in the Belgian Blue (Figure 4.13 (e)). Because this animal is a heterozygote, it is not expected to have a double muscled phenotype. However, the presence of a non-functional allele should cause an intermediate muscle phenotype, as discussed in Section 4.4.2. A more detailed analysis of the muscle fibres will be carried out at a later time to confirm the animal's phenotype as the current priority is to breed from this animal.

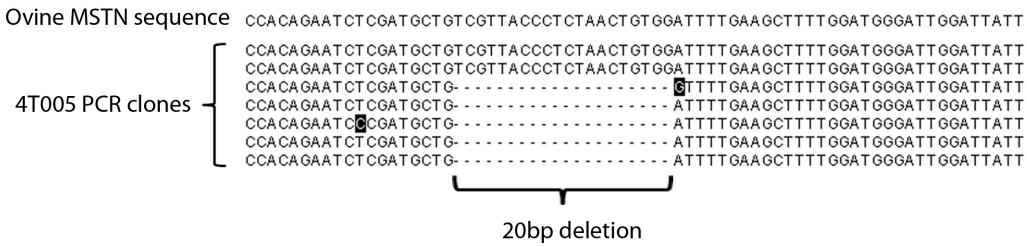
The restricted breeding season and small offspring numbers of sheep mean that it is often necessary to inject multiple editors into the same embryo, to maximise the use of limited resources. Because of this, the CRISPR-edited lamb had also been injected with sgRNAs targeted against PPT1 as part of a separate project in the group (PPT1 sgRNAs were designed and validated by Dr Simon Lillico). Sequencing was carried out to determine whether the PPT1 locus had also been edited in this animal. Figure 4.14 (a) shows a alignment between the PPT1 locus amplified from the edited lamb against the PPT1 sequence available in the sheep genome. No indications of editing are present, despite confirmation of the PPT1 sgRNAs' activity *in vitro* (Dr Simon Lillico, personal communication). The ovine myostatin cDNA sequence was also checked to ensure that it contained no off-target binding sites for the PPT1 sgRNA. The largest area of homology detected was a 6bp fragment in exon 2 of myostatin, with no appropriate PAM sequence present to allow cleavage by Cas9 (Figure 4.14(b)).

(a)

(b)



(c)



(d)

GGCTTGATTGTGATGAGCACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGAAGCTTTT

Belgian Blue deletion

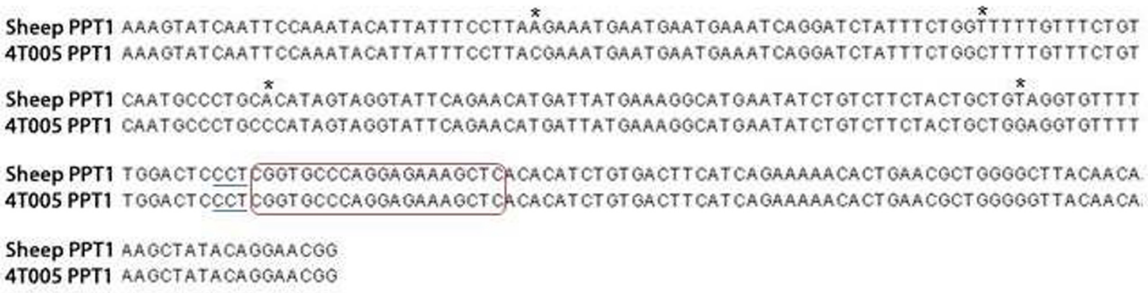
4T005 deletion

(e)

DFGLDCDEHSTESRCCRYPLTVDFEAFGWDWIIAPKRYKANYCSGECEFLFLQKYPHTHLVHQANPKGSAGPCCTP
TKMSPINMLYFNGKEQIIYGKIPGMVVDRCGCS.

Figure 4.13.: Identification and analysis of myostatin-edited lamb produced using CRISPRs. (a) Image showing genome-edited sheep on right (yellow mark on back, ID number is 4T005) (b) Sequencing chromatogram from 4T005 DNA at the MSTN target site. The green line indicates the beginning of the editing event. as both traces have an equal signal intensity, the animal was determined to be a heterozygote. (c) Alignment of sequences from PCR clones of 4T005 DNA against the ovine MSTN sequence. The 20bp deletion event is clearly visible. Other nucleotides which differ from the ovine sequence are highlighted (d) Section of the ovine MSTN sequence annotated to show the location of the deletion event in 4T005 (green) in relation to the 11bp Belgian Blue deletion (blue). Nucleotides which, in the presence of the 20bp deletion found in 4T005, form a TGA stop codon are highlighted in red. (e) Amino acid sequence of the mature myostatin protein. Amino acids which are absent in both Belgian Blue cattle and in one allele of 4T005 are coloured in orange. Residues which are present in Belgian Blue cattle but absent in 4T005 are coloured in red.

(a)



(b)

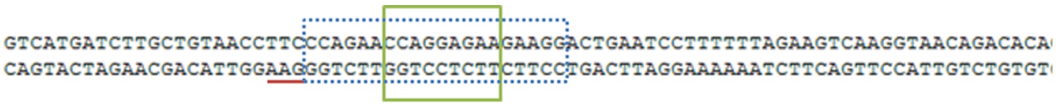


Figure 4.14.: Analysis of potential PPT1 editing in the CRISPR-edited lamb (ID - 4T005). (a) Sequence alignment of the PPT1 locus in 4T005 against the ovine genomic sequence. The binding site of the PPT1 sgRNA is shown by a red box and the associated PAM sequence underlined. SNPs between the two sequences are shown by asterisks. (b) Section of exon 2 of ovine myostatin showing area of homology with the PPT1 sgRNA binding site (green box). Based on this area of homology, the length of the whole sgRNA is shown by the dashed blue box and the site where the PAM should be is underlined.

4.3.5. Determining the presence of the Texel G/A SNP in the edited animals

As both edited lambs resulted from abattoir-derived oocytes, their genotype at other loci of the myostatin gene is unknown. The abattoir occasionally receives Texel sheep for slaughter, so there was a possibility that one or both of the edited lambs could carry the Texel SNP described in Section 4.1.3 and, therefore, could be compound heterozygotes. The location of the Texel SNP in the 3'UTR has already been determined (Takeda et al., 2010) and DNA from both animals was sequenced to see if the Texel SNP was present. Figure 4.15 shows the results of the sequencing. Both animals are homozygous for the wildtype G at the SNP, indicating that the Texel myostatin mutation is not present.

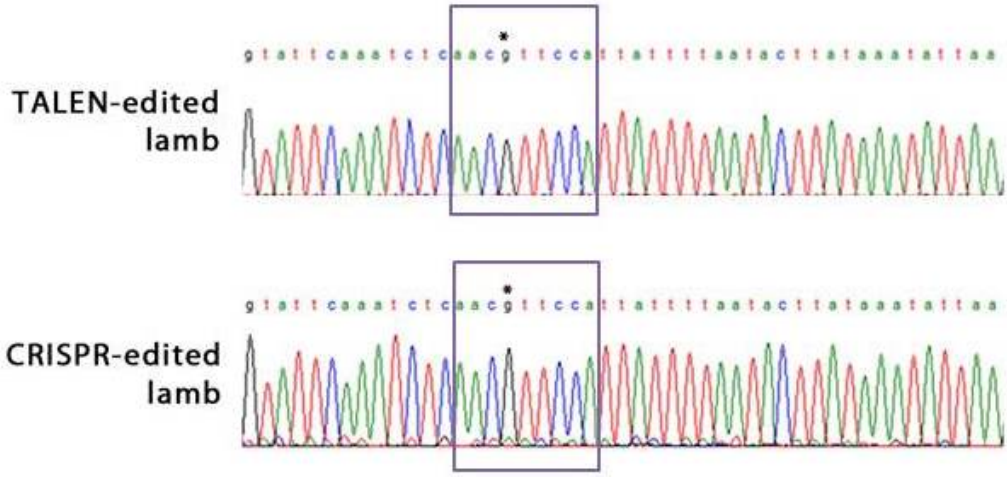


Figure 4.15.: Chromatograms showing that both myostatin-edited lambs are homozygous for the 'G' SNP in the 3'UTR of myostatin and not the Texel 'A' SNP, which creates an octomer motif (highlighted in purple box) and, as a result, an illegitimate miRNA binding site. The SNP is marked on each chromatogram by an asterisk.

4.4. Discussion

This chapter demonstrates how embryo microinjection of genome editors, specifically TALENs and CRISPRs, can be used to introduce a desired allele into a particular breed or species, as exemplified here with a non-functional myostatin allele. The speed with which an allele can be introduced into a population using genome editors as opposed to conventional breeding is extremely attractive for agricultural purposes. This is especially true for sheep, which only breed

once a year. For example, it took 30-40 years of breeding the Texel SNP into the Norwegian White sheep breed before the allele was almost fixed in the population (Boman et al., 2011). By having the ability to produce multiple animals containing the desired allele in one breeding season via embryo microinjection, the time an allele takes to reach fixation in a population can be drastically reduced.

Two edited animals were produced in this project, one from each type of genome editor used. They are discussed in greater detail below, with the validation of genome editors discussed in Chapter 6.

4.4.1. Effects of the edited myostatin allele in the TALEN-edited lamb

The TALEN-edited lamb was found to have a 3bp deletion at the myostatin target site. The low signal intensity of the edited sequence on a chromatogram indicated that the editing event had occurred after the injected zygote had undergone cell division and that only a proportion of the cells in the animal actually contained the edited allele. This means that any phenotypic effect exerted by the edited allele would potentially be reduced due to the predominance of the wildtype allele.

The deletion led to the loss of a single amino acid, arginine 283, from the mature myostatin protein. Interestingly, microinjections using the MSTN TALEN mRNA in bovine zygotes and carried out in parallel to this work obtained strikingly similar editing events in two out of the three edited cattle obtained in total (Proudfoot et al., 2014). As with the edited lamb, one of the edited cattle was mosaic, with two edited alleles in addition to a wildtype allele, again indicating that the TALENs were active after cell division. In addition, one of these edited alleles also had a deletion of arginine 283. A second edited calf had a myostatin allele where cysteine 282 was lost from the protein. It seems unlikely that the loss of a single amino acid from the protein would be capable of exerting a significant phenotypic effect, unless the amino acid played a key functional or structural role in the protein.

Analysis of the crystal structure of the mature myostatin protein complexed with the inhibitor follistatin (Cash et al., 2009) shows that arginine 283 is located within a β-strand and may participate in hydrogen bonding with aspartic acid 271. It is possible that loss of arginine 283 may cause a change in the structure of the myostatin protein either through the loss of this hydrogen bond or by the loss of the arginine residue impacting on the local structure, particularly the di-

sulphide bridge formed by cysteine 281. This could result in a hypomorphic myostatin allele, where the protein's ability to dimerise or to bind inhibitors such as the myostatin propeptide are altered (Cash et al., 2009, 2012), although this would need to be confirmed by modelling of the edited protein's structure and *in vitro* activity assays. However, the effects of this allele are unlikely to be as severe as those seen when a myostatin-null allele is present and may not result in a visible phenotype.

4.4.2. Effects of the edited myostatin allele in the CRISPR-edited lamb

The CRISPR-edited lamb was heterozygous for a 20bp deletion, which is predicted to result in one allele which produces a severely truncated myostatin protein, in addition to the WT allele. As the edited protein is predicted to be shorter than that found in Belgian Blue cattle, it is very likely that this allele is non-functional. Studies of animals heterozygous for a non-functional myostatin allele have shown these animals exhibit an intermediate phenotype, suggesting that myostatin has an additive effect. As an example, the muscle weights of heterozygous mice have been found to fall between those of wildtype and myostatin-null animals (McPherron and Lee, 2002) and increased muscularity has been observed in larger heterozygous animals, such as dogs (Mosher et al., 2007).

This intermediate phenotype is also seen in heterozygous livestock animals and can increase the agricultural value of the animals. Lambs which are heterozygous for the Texel G/A SNP showed an increased growth rate and improved feed conversion rate in comparison to wildtype controls (Haynes et al., 2012), while calves heterozygous for inactivating myostatin mutations had a heavier birth weight and a higher carcass score (i.e. increased muscle content and decreased fat content) at slaughter (Allais et al., 2010).

As mentioned previously, the main priority for future work is to use this animal for breeding in a n effort to expand the edited allele into a population. However, an application has also been made to the Home Office for approval to take muscle biopsies from this animal under anaesthesia for subsequent analysis. This should indicate whether the edited lamb does indeed exhibit the intermediate phenotype found in heterozygotes.

4.4.3. Inactivating myostatin at different developmental stages

As described in Section 4.1.3, knocking out myostatin during embryonic development results in increased muscle mass via an increase in both hypertrophy and hyperplasia. Several studies have investigated the effects of inactivating myostatin in adult animals, through either conditional deletion of myostatin or inhibiting myostatin using antibodies. Although an increase in muscle mass is seen when myostatin is inactivated in these animals, it is not as great as the increase seen when myostatin is knocked out during embryonic development. This is due to the fact that the increase in muscle mass is solely due to an increase in hypertrophy with no contribution from hyperplasia (Whittemore et al., 2003, Welle et al., 2007), as adult muscles exhibit a restricted ability to grow by hyperplasia (Levy et al., 2006). Therefore, inactivating myostatin by embryo microinjection of genome editors can produce greater enhancement of muscle mass in the resulting animal than any interventions which are taken when the animal is already mature as it is still possible to increase muscle growth by hyperplasia as well as hypertrophy.

4.4.4. Economic viability of the myostatin-knockout phenotype in sheep

Despite the clear agricultural value of the double-muscled phenotype in agriculture, knocking out myostatin in livestock also brings disadvantages, primarily with breeding. Belgian Blue calves are routinely delivered by Caesarian section as the increased muscularity of both the cow and calf mean that the calf is unable to travel through the birth canal unassisted (Bellinge et al., 2005). These birthing difficulties are a significant trade-off against the increased productivity of double-muscled livestock and may make these animals unsuitable in some circumstances. Sheep are a particularly good example of this, as they tend to be farmed less intensively and with less supervision than cattle. This makes the establishment of interventions such as delivery of lambs by Caesarian section more problematic as sheep tend to be more dispersed across farmland and need to be brought together. Further to this, sheep breeds such as the Herdwick, which are left to graze on poor-quality pasture for prolonged periods rather than being actively fed are likely to show fewer benefits from an myostatin-inactivating mutation as those which receive feed of a higher quality, which will have a greater growth rate. (Haynes et al., 2012).

These restrictions may be possible to circumnavigate through careful choice of the nature

of myostatin mutation used and the recipient breed. The Texel myostatin mutation is less severe than that of the Belgian Blue and, as a result, Texel ewes do not encounter birthing difficulties. It is possible that introduction of a similar knockdown mutation using genome editors instead of a knockout could bring the desired productivity advantages without causing birthing difficulties and is something that should be investigated further. An alternative method could be to use breeds which are known for giving birth with relative ease, such as the Nelore cattle (Proudfoot et al., 2014) as the recipients for these mutations, as they may be less affected by the complications caused by the double-muscled phenotype during labour.

5. Using genome editors to engineer host resistance to FMDV proteases

5.1. Introduction

Foot-and-mouth disease (FMD) is a highly contagious disease which primarily affects cloven-hoofed livestock animals, causing painful vesicles to form around the feet and mouth and leading to lameness and loss of productivity in the animal. Although the disease has a mortality rate of ~1%, it has a 100% morbidity rate and causes considerable production losses for farmers through weight loss or low milk production, particularly as highly productive livestock breeds are more susceptible to FMD (Parida, 2009, Knight-Jones and Rushton, 2013). The FMD epizootic in the UK in 2001 alone is estimated to have cost $12-20 billion in production losses and with 75% of the global population living in FMD-enzootic countries, such as those in sub-Saharan Africa, reducing the spread of FMD could bring significant economic benefits (Parida, 2009, Knight-Jones and Rushton, 2013).

The causative agent of FMD is foot-and-mouth-disease virus (FMDV), an RNA virus and member of the *Picornaviridae* family. Seven immunologically distinct FMDV serotypes have been identified based on the neutralising antigenic sites found on the capsid proteins, with numerous subtypes within each serotype. These serotypes, O, A, C, SAT 1, SAT 2, SAT 3 and Asia 1, along with the equine rhinitis A virus, make up the *Aphthovirus* genus and FMDV is considered to be the prototype aphthovirus (Fields et al., 2007, Mahy, 2005).

FMDV primarily infects members of the order *Arteriodactyla*, although more species can either be infected experimentally or can carry the disease without showing clinical symptoms (Thomson et al., 2003, Weaver et al., 2013). The virus can be transmitted in a number of direct and indirect ways, but generally infects a susceptible animal either by inhalation of aerosolised

virus or through open wounds in the skin. FMDV initially infects epithelial cells which express integrin receptors (described further in Section 5.1.3), such as those on the soft palette, tonsil and pharynx (Prato Murphy et al., 1999), before spreading through the body via the lymph nodes and circulatory system. The main sites of FMDV replication are epithelia in the skin, mouth, tongue and tonsils (Mahy, 2005, Stenfeldt et al., 2014a).

Although an FMDV infection can generally be cleared by the host in a number of days, different livestock species show significant differences in their susceptibility to the virus and their ability to transmit the virus to other animals. Pigs and cattle generally exhibit severe clinical symptoms upon FMDV infection, in contrast to sheep, which only show mild symptoms. However, pigs require a far greater dose of FMDV than sheep or cattle to become infected but, once infected, produce significantly more aerosolised virus than either sheep or cattle. Finally, there is evidence that both sheep and cattle can become carriers of FMDV once the clinical symptoms have subsided. Sheep are only thought to exist in this carrier state for a number of weeks but cattle have been found to be FMDV carriers over three years after infection. This is compounded by the fact that around 50% of infected cattle become FMDV carriers after infection, forming a significant reservoir for the virus. There is no evidence for a carrier state in pigs (Alexandersen et al., 2002, Kitching, 2002, Kitching and Alexandersen, 2002, Kitching and Hughes, 2002, Stenfeldt et al., 2014b).

Vaccines against FMDV have been developed using inactivated virus and form an integral part of some countries' FMDV containment measures, despite having a number of drawbacks. The use of inactivated virus carries a risk that some active viral particles may be present in the final vaccine, although recent research into the production of empty FMDV capsids could solve this issue (Porta et al., 2013). In addition, the vaccines can only provide protection against the specific serotypes and subtypes used to produce the vaccine and can be overcome by high levels of circulating virus (Kitching, 2002). The vaccines also cause problems for farmers in FMDV-enzootic areas as it is difficult to distinguish a vaccinated animal from one which has been infected with FMDV. This can restrict a farmer's trading opportunities as countries which are free of FMDV will only import goods from certified disease-free animals. This emphasises the need for improved FMDV control measures.

5.1.1. Possible uses of genome editors in FMDV control

The RNA genome of FMDV renders it unsuitable as a genome editing target, as the high mutation rate of the virus makes it difficult to implement long-term control measures which rely on the conservation of a particular structure or sequence which has been inserted into the genome using genome editors. This is particularly true if the inserted sequence reduced the fitness of the virus, which would be a likely aim of any FMDV control measures targeting the virus itself.

However, FMDV is known to interact with a number of host proteins at various points in its life cycle (see Figure 5.1). The most notable examples of these interactions are cleavage of host proteins by the FMDV proteases, as described in Section 5.1.4. These interactions are key successful replication of the virus and therefore present a range of attractive alternative targets for genome editing and FMDV control. If these host proteins could be edited so that the viral interactions are abolished, this could provide a way to introduce a degree of FMDV resistance into a livestock population. This chapter describes the selection of potential target host proteins and the design of genome editors to introduce specific mutations which may abolish interactions between these proteins and FMDV.

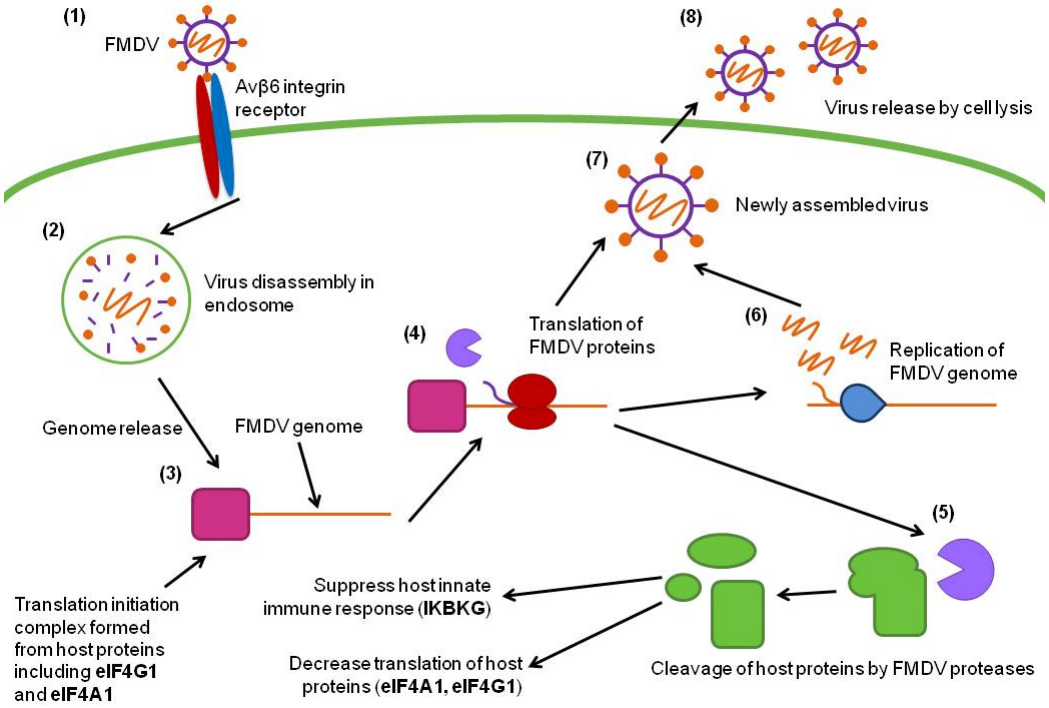


Figure 5.1.: The life cycle of FMDV. The three host proteins which are the subject of research in this chapter are shown in bold. (1) FMDV particles enter a cell via integrin receptors on the cell membrane. (2) Viral particles undergo disassembly in early endosome, allowing the FMDV gneome to be released into the cytoplasm. (3) A translation initiation complex forms at the 5'end of the FMDV genome. This complex consists of proteins form the host cell, including eIF4A1 and eIF4G1. (4) The FMDV genome is translated, allowing the L and 3C proteases to be produced. (5) The L and 3C proteases cleave host proteins. This can have numerous effects in the cell, including the shutoff of cellular translation and the innate immune response. (6) Copies of the FMDV genome are produced by the virus' 3D polymerase. (7) New viral particles are assembled and released (8) by cell lysis.

5.1.2. The structure of FMDV

FMDV is a spherical, non-enveloped virus with a diameter of about 30nm (Acharya et al., 1989). The icosahedral virus capsid is made up of four structural proteins, VP1, VP2, VP3 and VP4, which encapsulate a positive sense ssRNA genome found on a single molecule (Fields et al., 2007). The genome is approximately 8.3kb long and contains a single ORF of about 7kb. The protein VPg is covalently linked to the 5' end of the FMDV genome and is removed upon cell entry to act as a primer for genome replication (Mahy, 2005). The FMDV genome attached to VPg is an infectious agent as the presence of this alone in a host cell is sufficient to trigger viral replication (Belsham and Bostock, 1988).

Translation of the ORF produces a polyprotein which is processed by viral proteases to produce the mature viral proteins. In addition to the four structural proteins, the polypeptide also produces non-structural proteins including the leader (L) and 3C proteases (discussed in greater detail in Section 5.1.4) and the RNA-dependent RNA polymerase 3D. Translation of the polyprotein is mediated by a Type II IRES at the 5' end of the genome (Chamond et al., 2014) and can be initiated from one of two AUG codons located at the 3' end of the IRES (Belsham and Brangwyn, 1990), although only the second AUG after the IRES is absolutely necessary for genome translation (Cao et al., 1995). The FMDV IRES has extensive secondary and tertiary structure, which is recognised by the host translation initiation factors (Lozano et al., 2014).

5.1.3. FMDV replication

As mentioned previously, FMDV enters cells via integrin receptors at the plasma membrane, which recognise the RGD sequence found on the βG-βH loop of the capsid protein VP1 (Fox et al., 1989, Leippert et al., 1997). Integrins are heterodimers consisting of one α and one β subunit and FMDV can be bound by integrins consisting of an αv subunit complexed with a β1, β3, β6 or β8 subunit (Jackson et al., 1997, Neff et al., 1998, Jackson et al., 2000a,b, 2004). The αvβ6 receptor is the predominant receptor on epithelial cells infected by FMDV and, as such, is considered the main determinant of virus tropism *in vivo* (Monaghan et al., 2005). Different FMDV serotypes exhibit different integrin preferences, which are thought to be influenced by the amino acids surrounding the RGD binding sequence (Jackson et al., 1997) and some serotypes have also been found to use heparin sulfate as a receptor (Fry et al., 1999). As FMDV is capable of binding to integrins from species which are not considered natural hosts of FMDV, it is unlikely that these molecules determine the host range of the virus (Jackson et al., 1997). Ruiz-Saenz et al. (2009) provide more details on the receptors used by FMDV.

Upon binding of integrin receptors, FMDV is internalised into the host cell by clatherin-dependent endocytosis and localised to early-stage or recycling endosomes (Berryman et al., 2005). As FMDV is very labile, the low pH environment of the endosome causes disassembly of the capsid (Carrillo et al., 1984, 1985) and release of the genome into the cytoplasm for subsequent translation and replication (Belsham and Bostock, 1988). The process by which

the FMDV genome is released from the endosome is currently unknown and is an area of active research, particularly at the Pirbright Institute.

Although the FMDV genome is uncapped, it is translated using components of the host cap-dependent translational machinery, which is subverted by the virus. Initiation of translation requires the presence of the host factors Met-tRNA$_i$, eIF2, eIF3, eIF4A, eIF4B, eIF4E, eIF4G and the ribosomal 40S subunit (Pestova et al., 1996a, Pilipenko et al., 2000). The initiation complex is stabilised by PTB (Kolupaeva et al., 1996) and the ribosomal 60S subunit is recruited to the FMDV genome after this initiation complex has formed. As translation proceeds, eIF2 becomes dispensable and is potentially degraded (Welnowska et al., 2011, Redondo et al., 2011) and PTB and eIF3 are cleaved by the FMDV proteases (Rodriguez Pulido et al., 2007). It is, as yet, not clear whether this degradation of key translation initiation factors is a mechanism of switching the FMDV RNA from translation to genome replication, thereby ensuring that the translating ribosomes do not disrupt progression of the 3D polymerase along the genome, or if degradation of these factors is just part of the 'collateral damage' incurred by the presence of the active FMDV proteases in the host cell.

Replication of the RNA genome is carried out by the 3D polymerase using VPg as a primer. A negative-sense RNA intermediate is first synthesised, from which many copies of the positive-sense genome can be produced and encapsulated in the newly synthesised capsid proteins. New virus particles are released upon lysis of the host cell (Fields et al., 2007).

5.1.4. The FMDV proteases and their interactions with host proteins

The two FMDV proteases, L and 3C, are responsible for the majority of processing of the FMDV polyprotein and for the subversion or destruction of host proteins and pathways, as reviewed by Chase and Semler (2012). Preventing protease cleavage of host proteins and maintaining host functions, including immune responses to the presence of the virus, presents an attractive mechanism for reducing FMDV replication in the host cell.

Both proteases are described below, with their targets discussed at later points in this chapter.

5.1.4.1. L protease

The L protease is a cysteine protease which is only found in aphthoviruses and cardioviruses and is structurally similar to the papain protease found in many different organisms, from vir-

uses to plants and animals (Guarne et al., 1998, 2000). Its catalytic activity relies on the amino acids C51 and H148 (Roberts and Belsham, 1995) and its cleavage specificity, although not yet completely elucidated, is in part mediated by the leucine or methionine residue at position 143 (Mayer et al., 2008). The position of L protease at the 5'end of the FMDV ORF means that two forms of the enzyme, named Lab and Lb, are produced from the two AUG codons at the start of the ORF. While both forms of the protease have the same activity *in vitro* (Medina et al., 1993), Lb is preferentially expressed *in vivo* (Cao et al., 1995).

The L protease catalyses its own release from the polyprotein (Strebel and Beck, 1986) and maintains translation from the IRES in the later stages of infection, when eIF2 is degraded (Moral-Lopez et al., 2014). In addition to its protease activity on both the FMDV polyprotein and on host proteins, the L protease can act as a viral deubiquitinase. This activity is central to the L protease's role in disrupting innate immune signalling pathways such as the Type I IFN response, where ubiquitin is used as an activator for many of the components in the pathway, such as Rig-I (Wang et al., 2011b).

5.1.4.2. 3C protease

The crystal structure of the serine protease 3C shows its structural similarities to chymotrypsin, including its active site, which is contained within a peptide-binding groove (Zunszain et al., 2010) and holds the 3C catalytic triad of H46, D84 and C163 (Grubman et al., 1995, Birtley et al., 2005). 3C cleaves at sites which have either a glutamic acid or glutamine residue at the P1 position (Zunszain et al., 2010) and carries out the majority of processing events on the FMDV polyprotein (Bablanian and Grubman, 1993). Like the L protease, the 3C protease also plays a role in blocking the host Type I IFN response. It blocks the nuclear translocation of Stat1/Stat2, reducing the activity of interferon-stimulated regulatory elements (ISREs) and therefore, the expression of interferon-stimulated genes (ISGs) (Du et al., 2014).

5.1.5. Selection of target sites

In this project, genome editors were designed to target host genes whose protein products are cleaved by FMDV proteases, with an eventual aim of engineering protease-resistance into the FMDV cleavage sites via the introduction of exogenous sequence into the genes by HDR. Several host genes known to be cleaved by FMDV proteases are to be found in the literature,

including eIF4G2 (Gradi et al., 2004). However, in some cases, the information given was either not complete enough to allow for the design of genome editors or did not correspond with the current genome and protein data for a particular species. This limited the number of genes which could be chosen as suitable targets. The three genes selected to be targeted were IKBKG, eIF4G1 and eIF4A1. Further details about each target gene's protease sites is given in Table 5.1 below;

Gene	Location of cleavage site	Protease	Reference
IKBKG	Between Q383 and R384	3C	Wang et al. (2012a)
eIF4A1	Between E143 and V144	3C	Li et al. (2001)
eIF4G1	Between G674 and R675	L	Kirchweger et al. (1994) Strong and Belsham (2004)

Table 5.1.: Table showing the location of protease cleavage sites in each of the selected host target proteins and the FMDV protease which targets each site

Given the key roles eIF4A1, eIF4G1 and IKBKG play in translation initiation and the innate immune response respectively, it is unlikely that reliance on modification of the genes by NHEJ alone would be able to produce protease-resistant alleles which maintained normal activity levels. Therefore, once genome editors which can successfully cleave at the target sites are developed, suitable templates for introgression by HDR would need to be designed. Previously published work has identified point mutations which can block cleavage of these host proteins by the FMDV proteases and could act as the basis for designing suitable HDR templates to introduce precise alterations. These mutations are described within each protein's section below.

5.1.5.1. IKBKG

IKBKG (also known as Nemo or IKKγ) is the regulatory subunit of the IKK complex in the NF-$\varkappa$B pathway. The IKK complex also includes the kinase subunits IKKα and IKKβ and activates NF-$\varkappa$B through the phosphorylation of I$\varkappa$B proteins. More detail on the role of the IKK complex in NF-$\varkappa$B signalling can be found in Israel (2010).

Cleavage of IKBKG by the 3C protease removes the C-terminal zinc finger domain which activates the full-length protein. Loss of this domain blocks signalling from the pattern recognition receptor (PRR) Mda-5, which detects the presence of picornaviruses in the host cell (Kato et al., 2006, Huesser et al., 2011), to both the NF-$\varkappa$B and IRF signalling pathways, resulting

in the loss of IFNβ expression (Wang et al., 2012a). This effect is also seen when IKBKG is cleaved by the 3C protease of another picornavirus, hepatitis A, although the cleavage site in this case is different to that of FMDV 3C (Wang et al., 2014a). Perturbation of the host innate immune response is a common feature in picornavirus infection and the interactions between the virus and host immune system have been recently reviewed by Feng et al. (2014).

Work identifying the 3C cleavage site in IKBKG also showed that mutation of glutamine 383 to an arginine residue could block cleavage of the protein *in vitro* (Wang et al., 2012a).

5.1.5.2. eIF4A1

The DEAD box RNA helicase eIF4A is essential for the successful initiation of translation from both capped and uncapped mRNAs as it forms part of the cap-binding complex eIF4F, along with eIF4G and eIF4E. The combination of eIF4A's RNA-dependent ATPase activity, in addition to its helicase function, allow it to unwind secondary structures found in the 5'UTRs of transcripts, preparing the RNA strand for ribosome binding (Ray et al., 1985, Pause and Sonenberg, 1992). There are three eIF4A isoforms; eIF4A1 and eIF4A2 show around 91% sequence homology at the protein level and are considered to be functionally equivalent (Nielsen and Trachsel, 1988), while eIF4A3 appears to play an inhibitory role in translation (Li et al., 1999). eIF4A1 is the most abundant eIF4A isoform found in the cell (Li et al., 1999).

In the early stages of FMDV infection, the FMDV IRES is recognised and bound by a complex of eIF4A1 and eIF4G1 (Pilipenko et al., 2000). Work in encephalomyocarditis virus, another picornavirus, has shown that this binding stimulates a conformational change in the IRES, probably as a result of eIF4A1's helicase activity (Kolupaeva et al., 2003).

Despite its importance in initiating translation from the FMDV genome, eIF4A1 is cleaved by the 3C protease during the course of infection. This cleavage inactivates the protein by removing an ATPase domain at the N-terminus, causing levels of both host and FMDV translation to decrease (Belsham et al., 2000, Li et al., 2001). As with other translation initiation factors mentioned previously, it is unclear whether cleavage of eIF4A1 used for FMDV replication is an unavoidable consequence of the virus shutting down cap-dependent translation or if it is of functional benefit to viral replication, as the removal of translating ribosomes allows unhindered replication of the FMDV genome by the virus' 3D polymerase. (Mahy, 2005). The eIF4A1 double mutant A142N/V144M has been found to be resistant to cleavage by the 3C

protease (Li et al., 2001).

5.1.5.3. eIF4G1

The two eIF4G isoforms, eIF4G1 and eIF4G2 share 46% identity in their amino acid sequences and have been shown to be functionally interchangeable in their role as an adaptor protein which binds the 40S ribosomal subunit (Pestova et al., 1996a, Hentze, 1997, Gradi et al., 1998). eIF4G1 is the predominant isoform in eukaryotic cells (Gradi et al., 1998). FMDV can only abolish cap-dependent translation in the host cell by cleaving both eIF4G isoforms (Gradi et al., 2004).

eIF4G1 is cleaved by the FMDV L protease (Devaney et al., 1988) and can be cleaved before the protease has self-processed from the FMDV polyprotein (Glaser et al., 2001). This cleavage reaction occurs very rapidly after virus entry; experiments using an RRL system, which is used to carry out cell-free translation of RNA molecules, show that the L protease can be detected four minutes after the addition of FMDV genomic RNA to the system, while complete cleavage of the RRL system's eIF4G1 can be seen eight minutes after the genome is added (Glaser and Skern, 2000). Work on other picornaviruses has suggested that this cleavage event may have an increased efficiency when eIF4G1 is bound to eIF4E, with the binding possibly inducing a conformational change, making eIF4G1 more favourable to cleavage (Haghighat et al., 1996). The C-terminal residues of the L protease, while not directly involved in protein cleavage, have been shown to be important for mediating this reaction (Glaser et al., 2001).

Cleavage of eIF4G1 removes the protein's N-terminal eIF4E binding site, which is essential for cap-dependent protein synthesis but not for translation from the FMDV IRES (Mader et al., 1995, Morino et al., 2000), and a NLS (Coldwell et al., 2004). This leaves the C-terminal fragment, which can bind eIF4A and eIF3, for use in FMDV replication (Ohlmann et al., 1997). As with cap-dependent translation, the eIF4G1 fragment is necessary to bind to the 40S ribosomal subunit to the FMDV genome (Pestova et al., 1996b).

Exchange of the L protease cleavage site in eIF4G1 for the analogous sequence of eIF4G2 (Foeger et al., 2002) can inhibit the protease's activity however, it remains unclear whether this also affects binding of the protease to eIF4G1. If binding is not prevented, the effects of L protease:eIF4G1 complexes on both viral replication and host translation will need to be evaluated. If necessary, the mutation of three additional amino acids K643A, K646A and

R650A, which have been shown to prevent the binding of L protease (Foeger et al., 2005), may be able to produce a fully resistant eIF4G1 allele.

Strong and Belsham (2004) demonstrated that eIF4G1 can also be cleaved by the 3C protease at E712/T713, downstream of the L protease site. Cleavage by 3C occurs after eIF4G1 has been cleaved by the L protease. However, eIF4G1's sensitivity to this cleavage event varies between species, with human eIF4G1 being resistant to the 3C protease, as it has a proline at position 713 instead of a threonine. It is also important to note that the majority of this work was carried out in BHK cells, and that the ovine, bovine and porcine eIF4G1 protein sequences share the P713 found in the cleavage-resistant human sequence. Therefore, it seems unlikely that this observation has biological relevance to the livestock species which commonly suffer from FMD.

5.2. Additional materials and methods

5.2.1. Bioinformatics

In order to find new potential targets of the FMDV proteases, the bovine and porcine proteomes were analysed using NetPicoRNA 1.0 (`http://www.cbs.dtu.dk/services/NetPicoRNA/`) (Blom et al., 1996). The bovine proteome used was release 2014_06 from Uniprot's Knowledgebase and contained a total of 24207 proteins, while the porcine proteome consisted of the 26143 proteins found in Uniprot's 2014_09 release. The ovine proteome was not analysed due to its relatively poor annotation.

Unfortunately, NetPicoRNA does not have an algorithm for predicting potential targets of the L protease, presumably due to the relative scarcity of literature available for predicting suitable L protease cleavage sites. Therefore, potential targets of the 3C protease were identified using the aphthovirus 3C algorithm. Each proteome was analysed in batches of 2500 sequences. Analysed sequences were given a cleavage score (i.e. how likely the sequence is to be recognised and cleaved by 3C) and a surface score (showing how exposed that particular sequence is on the protein) by NetPicoRNA. Scores range from 0 to 1. Candidates with both a cleavage score and surface score greater than 0.5 were designated as potential cleavage candidates by the software.

Following the analysis, any sequences not meeting the following criteria were removed from

the data set to ease further analysis;

1. Sequence should been designated a 'potential' target by Net PicoRNA
2. Sequence should have both a cleavage score and surface score above 0.75
3. Sequence should have either a glutamine or glutamic acid residue at the P1 position

Target predictions and data set cleanup were performed by Dr Zen Lu.

All uncharacterised proteins were then removed from each data set and the remaining candidates sorted in descending order of cleavage score and surface score. Gene clustering was also carried out on the final list of potential targets using PANTHER `http://www.pantherdb.org` (Mi et al., 2013) as well as the original bovine and porcine proteomes to determine whether any particular molecular functions or biological processes were enriched for in the final dataset. Additional functional enrichment analysis was performed using ToppFun (Chen et al., 2009a) with the p-value cutoff set at 0.05 and results grouped by the Biological Process, Molecular Function and Pathway features.

5.3. Results

5.3.1. Comparison of target sequences from different livestock breeds

There is some evidence in the existing literature that certain breeds or individual animals can exhibit some degree of resistance to FMDV infection (see Lei et al. (2012) and cited in Morris (2007)). To discover if any specific bovine, ovine or porcine breeds carried polymorphisms around the three protease cleavage sites of interest, amino acid and mRNA sequences from a selection of breeds were aligned. The amino acid sequences are shown in Figures 5.2, 5.3 and 5.4, while the mRNA sequence alignments can be found in Appendix A.5.

Although some amino acid substitutions were observed in eIF4A1, they are not situated in close enough proximity to the cleavage site to be considered to have a direct effect upon cleavage by the FMDV proteases. This does not exclude the possibility that the polymorphisms may exert a more indirect effect on protease cleavage (e.g. by affecting binding of the enzyme), but more information on the nature of the binding between the protease and these targets is required to determine this.

The alignment of eIF4G1 sequences from different sheep breeds show a 6 amino acid insertion/deletion between the consensus sequence, derived from the Ensembl database, and the breed-derived sequences. As the sequence from the Texel breed is included among the breeds analysed and the sheep genome is based on the Texel breed, it is likely that this is a result of an inaccuracy in the consensus sequence rather than a biological feature.

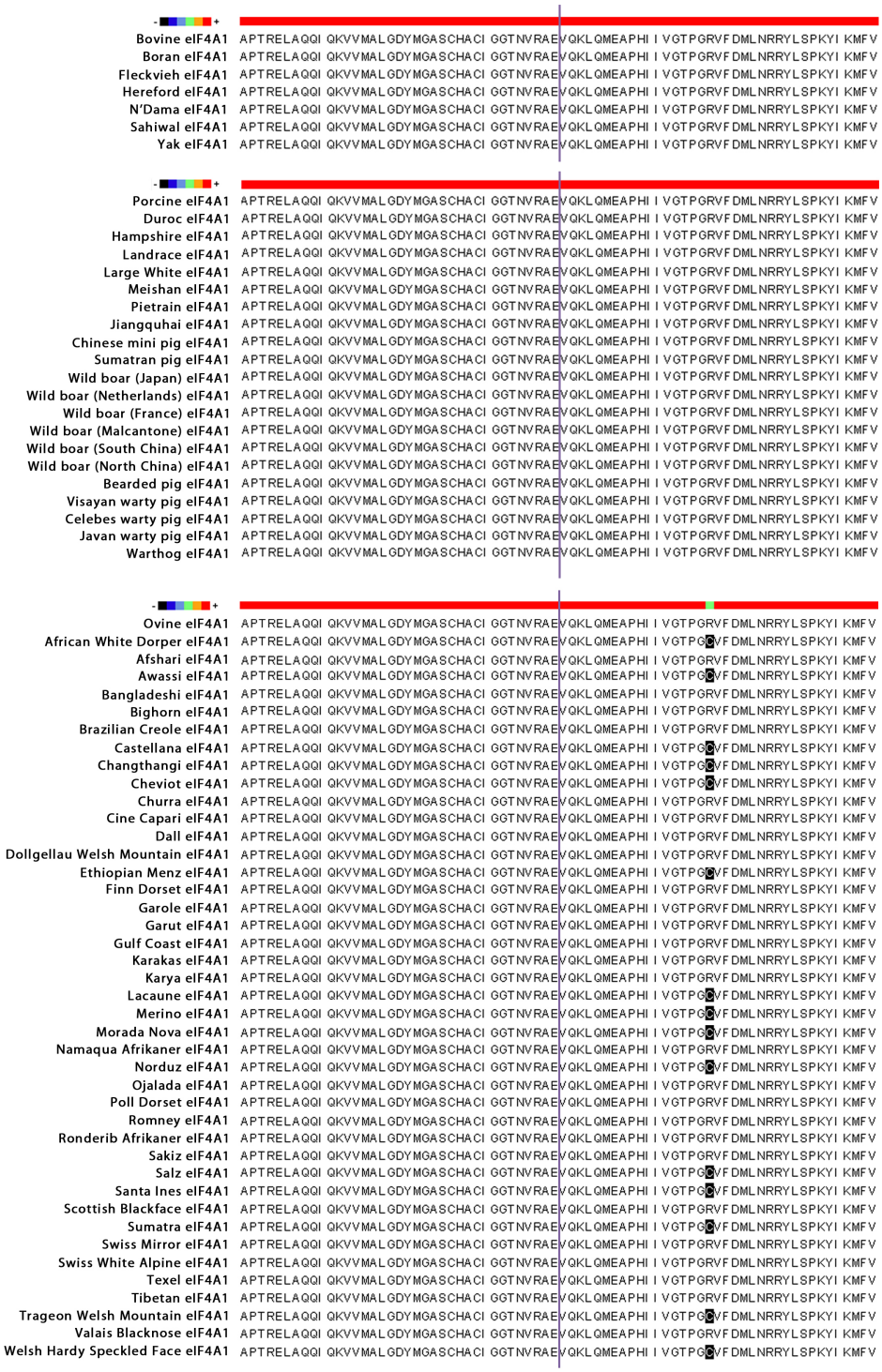


Figure 5.2.: Peptide alignments of eIF4A1 protein sequences between different bovine, porcine and ovine breeds. The bar across the top of each set of alignments shows the degree of conservation between the different sequences, with red indicating that all the sequences share the same amino acid residue at that particular point. Residues which differ from the consensus sequence are shown in a black box. The reference sequence from the published genome of each species is the top sequence of each alignment (Bovine eIF4A1, Porcine eIF4A1 and Ovine eIF4A1). The 3C protease cleavage site is indicated by the purple line.

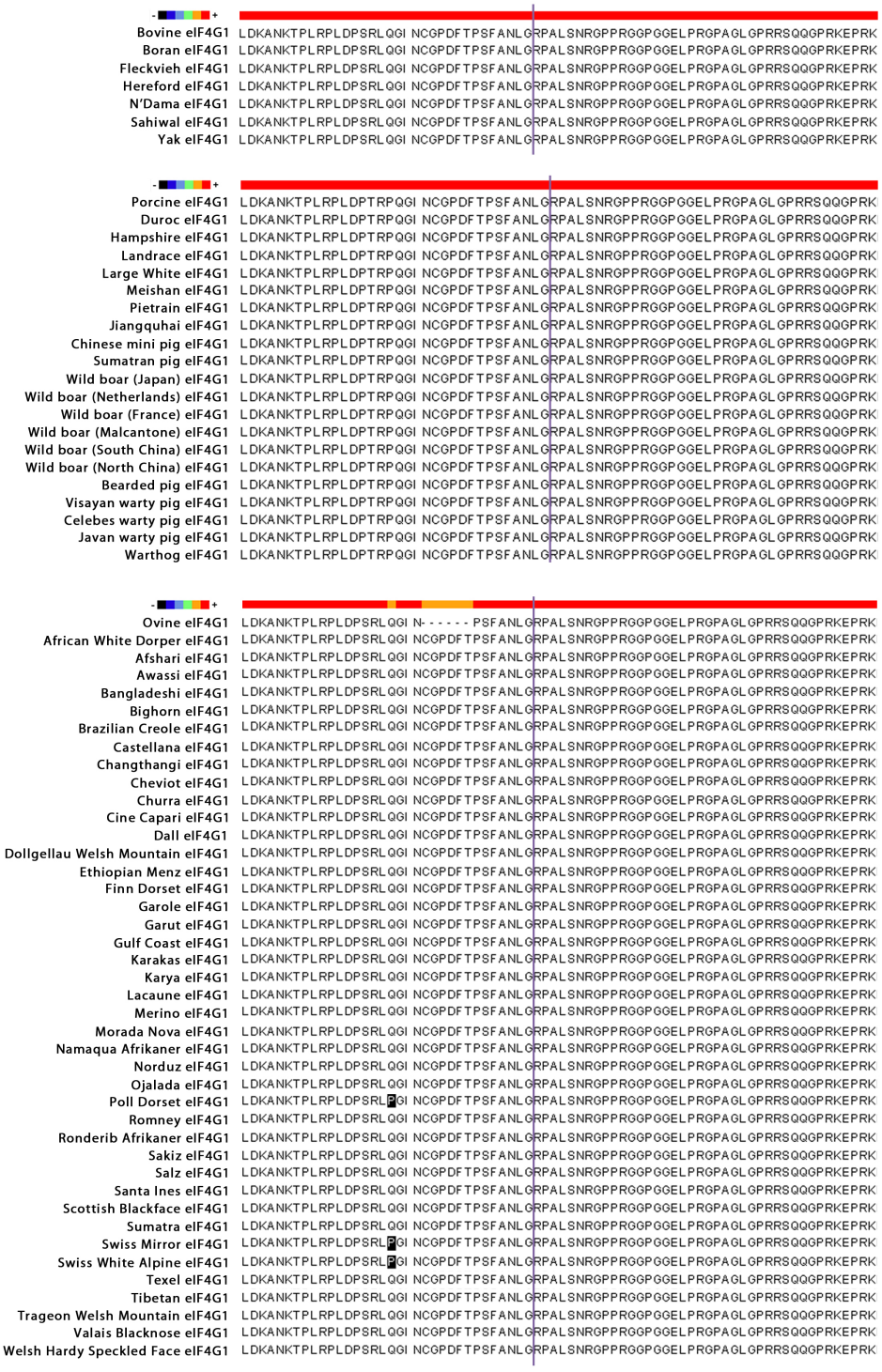


Figure 5.3.: Peptide alignments of eIF4G1 protein sequences between different bovine, porcine and ovine breeds. The bar across the top of each set of alignments shows the degree of conservation between the different sequences, with red indicating that all the sequences share the same amino acid residue at that particular point. Residues which differ from the consensus sequence are shown in a black box, with the exception of the 6aa insertion/deletion between the consensus and breed sequences. The reference sequence from the published genome of each species is the top sequence of each alignment (Bovine eIF4G1, Porcine eIF4G1 and Ovine eIF4G1). The L protease cleavage site is indicated by the purple line.

Figure 5.4.: Peptide alignments of IKBKG protein sequences between different bovine, porcine and ovine breeds. The bar across the top of each set of alignments shows the degree of conservation between the different sequences, with red indicating that all the sequences share the same amino acid residue at that particular point. Residues which differ from the consensus sequence are shown in a black box. The reference sequence from the published genome of each species is the top sequence of each alignment (Bovine IKBKG, Porcine IKBKG and Ovine IKBKG). The 3C protease cleavage site is indicated by the purple line.

5.3.2. Validation of TALENs

TALENs were designed to target the protease cleavage sites eIF4A1, eIF4G1 and IKBKG in the porcine (Figure 5.5) and ovine (Figure 5.6) genomes and constructed as detailed in Section 2.2. Three pairs of TALENs were designed for each gene in each species. Genomic sequences from all three target genes were obtained from the Ensembl database (Version 70). The intention was to initially develop active TALENs which would cleave at the protease sites and subsequently use these TALENs to introduce modified DNA sequences to the genes by HDR. The modifications would be designed so that the protease cleavage sites became protease-resistant while the endogenous activity of the proteins themselves remained unchanged. Validation of TALENs was carried out by co-transfection of the TALENs and a GFP plasmid into PK15 cells or OEFs as appropriate followed by a T7 assay.

Similar to results in previous chapters, no discernible TALEN activity was seen in any of the transfected porcine (Figure 5.8) and ovine (Figure 5.9) cells, despite the presence of GFP expression suggesting that the transfections had been successful. To confirm that the target sequence for each gene was conserved between the published consensus sequence and the cells used in the validation assay, PCR products were sequenced and, in most cases, showed that the target sites aligned with the sequences used to design the TALENs (Figure 5.10).

The exception to this was the ovine IKBKG B left TALEN binding site. This locus was disrupted by an ~200bp intron which had been absent from the genomic sequence originally used to design the TALENs (Figure 5.7), but was subsequently added to a newer version of Ensembl. This meant that the TALEN pair was not able to function as the left TALEN could not bind to a suitable sequence and allow the FokI nucleases to dimerise. A shortage of time and the numerous problems encountered in attempting to produce active TALENs meant that this TALEN was not redesigned.

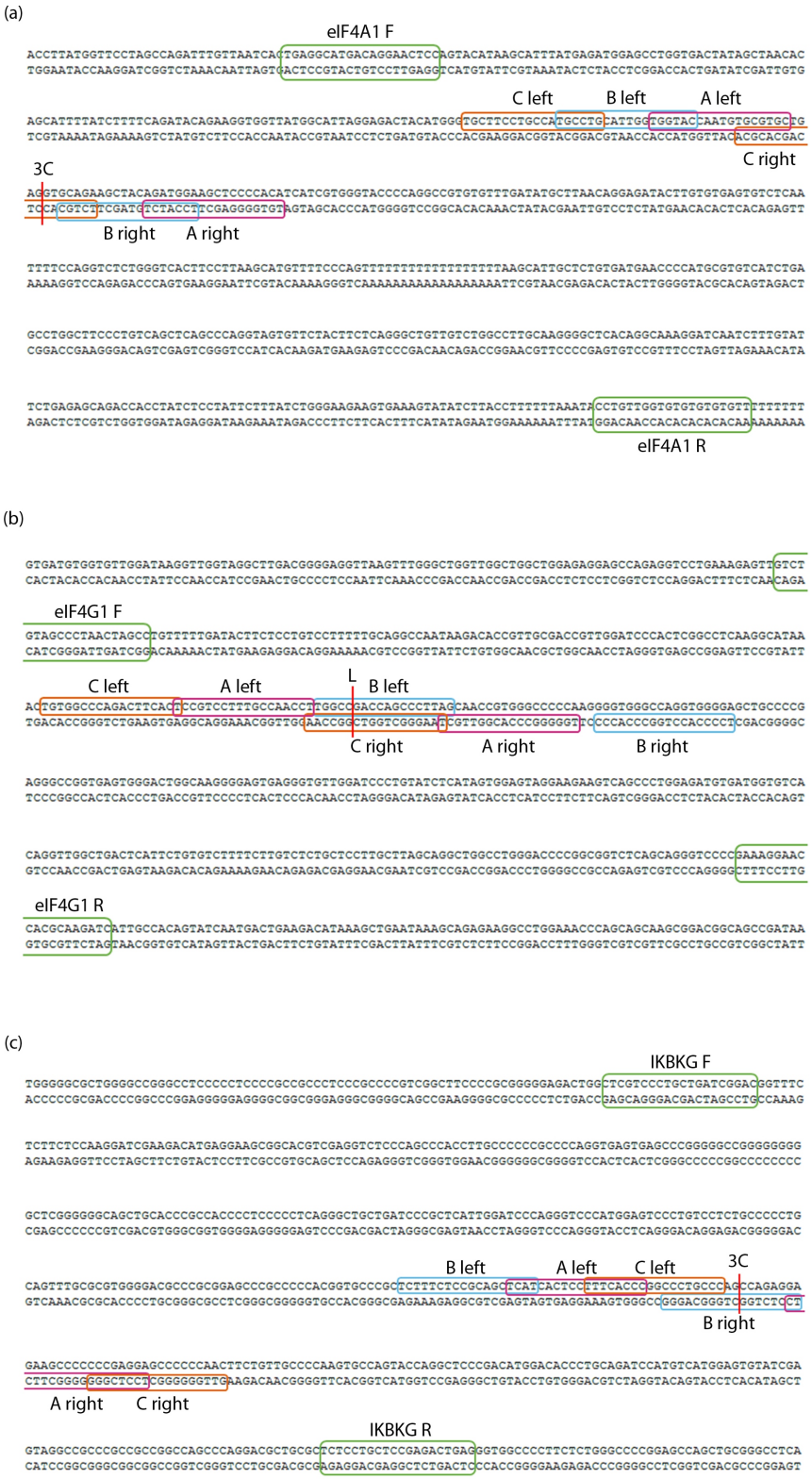


Figure 5.5.: Positions of TALEN binding sites in the porcine FMDV target genes (a) eIF4A1, (b) eIF4G1 and (c) IKBKG. Primers used for analysis are shown by green boxes and the protease cleavage sites are indicated by red lines.

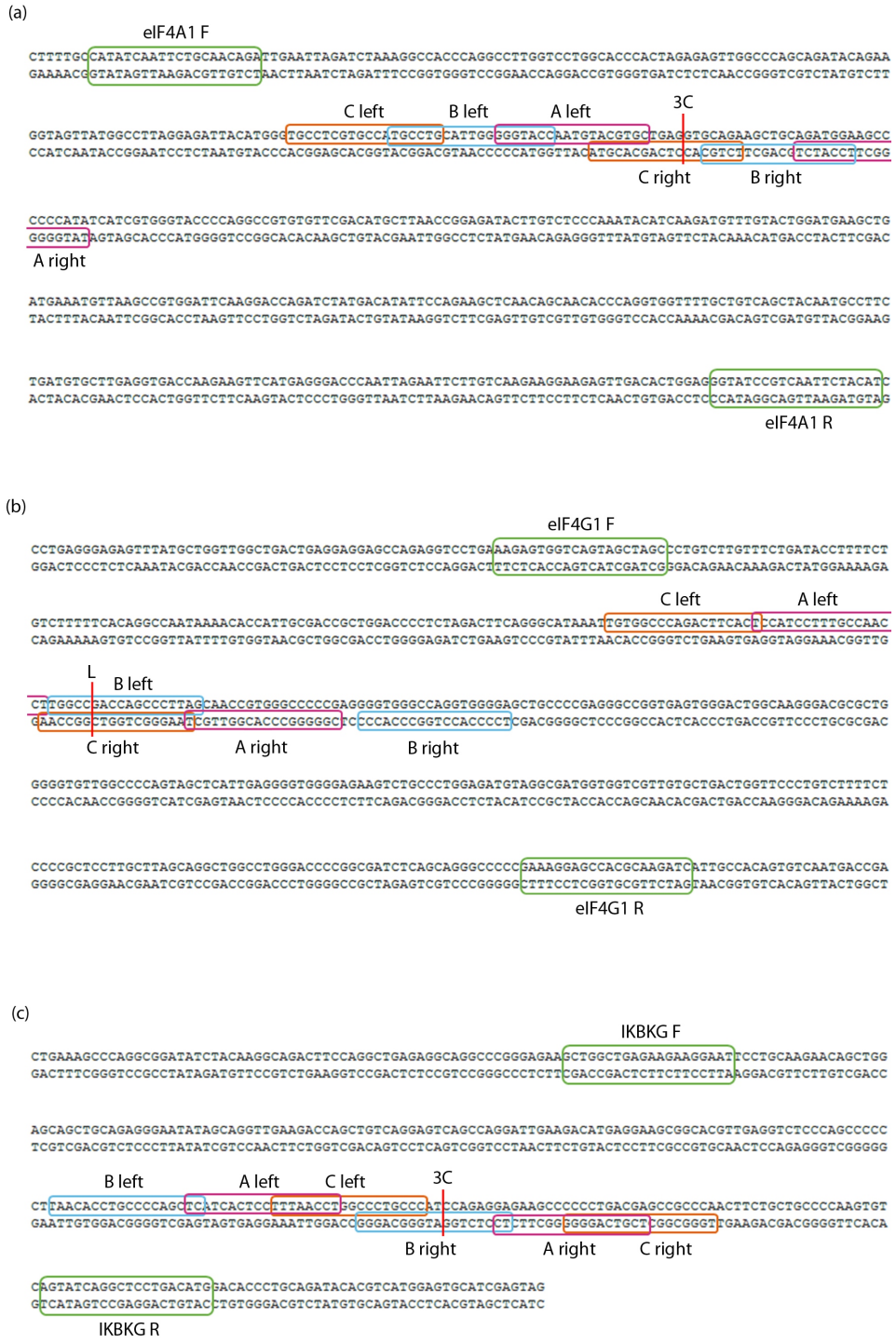


Figure 5.6.: Positions of TALEN binding sites in the ovine FMDV target genes (a) eIF4A1, (b) eIF4G1 and (c) IKBKG. Primers used for analysis are shown by green boxes and the protease cleavage sites are indicated by red lines. Note that this figure shows the genomic sequence for IKBKG taken from version 70 of Ensembl and not the updated sequence.

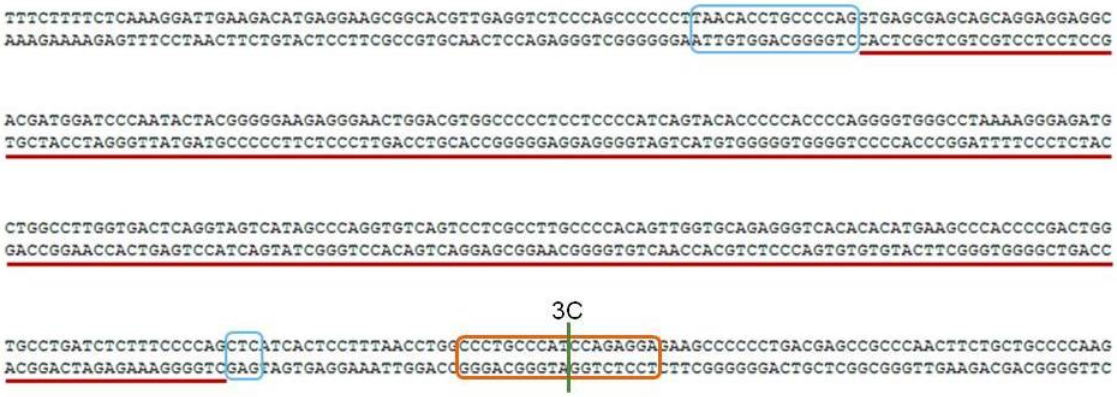


Figure 5.7.: Figure showing the disruption of the ovine IKBKG TALEN B left binding site by a previously unidentified intron. The nucleotides originally intended to be targeted by the B left TALEN are marked with blue boxes, while the intronic sequence is underlined in red. The IKBKG B right TALEN binding site is shown in the orange box, while the 3C cleavage site is indicated by the green line.

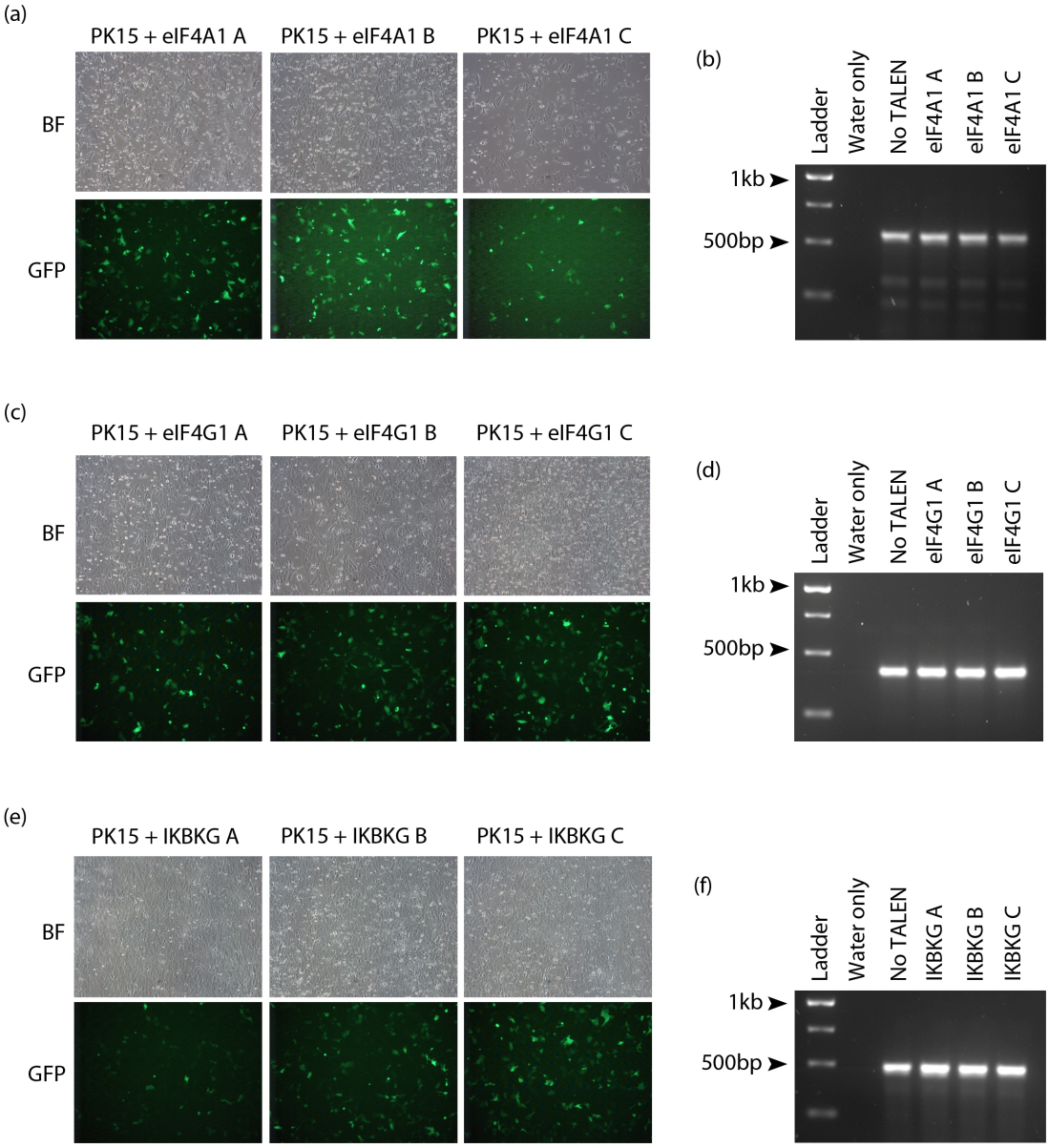


Figure 5.8.: Transfection of eIF4A1, eIF4G1 and IKBKG TALENs into PK15 cells. In each case, the TALEN pair was co-transfected with a GFP plasmid to indicate transfection efficiency. Brightfield and GFP images are shown for transfections of (a) eIF4A1 TALENs, (c) eIF4G1 TALENs and (e) IKBKG TALENs. Images at are 100x magnification. Gel images showing the results of T7 endonuclease digests are shown for (b) eIF4A1, (d) eIF4G1 and (f) IKBKG. Note that no detectable TALEN activity was seen in any of the gels.

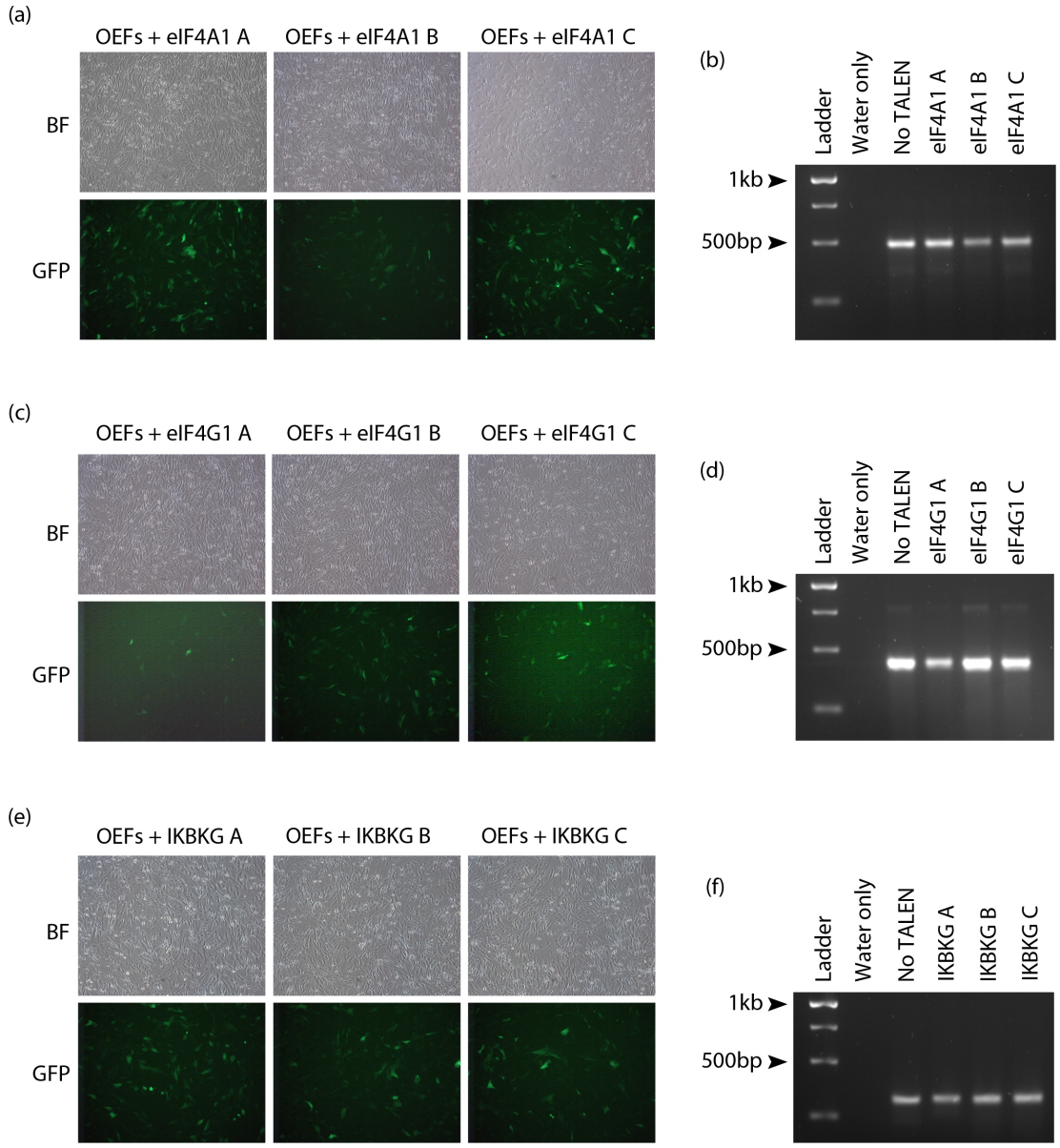


Figure 5.9.: Transfection of eIF4A1, eIF4G1 and IKBKG TALENs into OEFs. In each case, the TALEN pair was co-transfected with a GFP plasmid to indicate transfection efficiency. Brightfield and GFP images are shown for transfections of (a) eIF4A1 TALENs, (c) eIF4G1 TALENs and (e) IKBKG TALENs. Images at are 100x magnification. Gel images showing the results of T7 endonuclease digests are shown for (b) eIF4A1, (d) eIF4G1 and (f) IKBKG. Note that no detectable TALEN activity was seen in any of the gels.

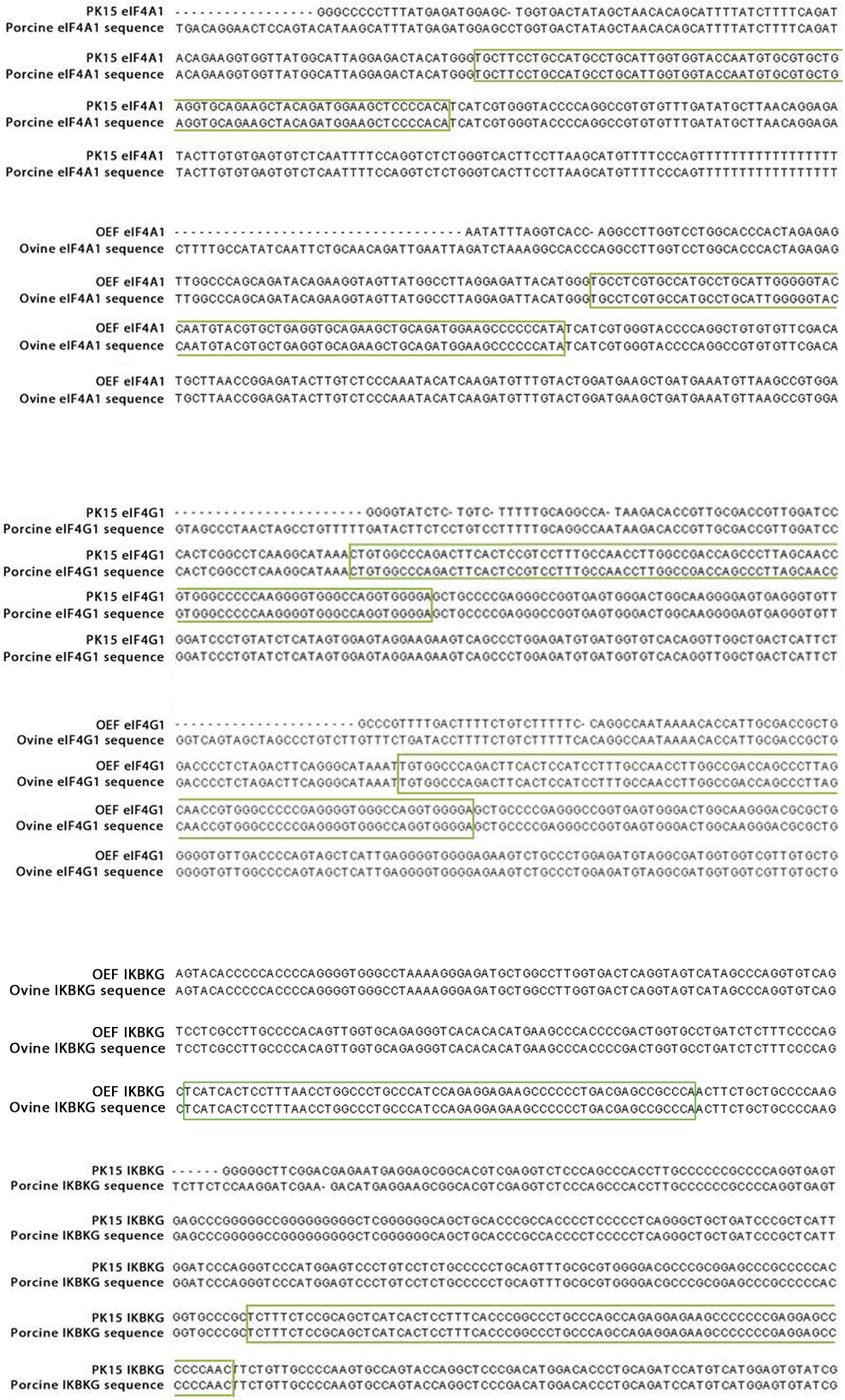


Figure 5.10.: Sequence alignments of TALEN target genes sequenced from OEF or PK15 DNA aligned against their counterparts in the ovine or porcine genome. Areas targeted by TALENs are highlighted in green boxes.

5.3.3. Bioinformatic prediction of 3C protease target genes

In order to predict other potential targets of the FMDV 3C protease and find additional candidates for future genome editing, bioinfomatic analysis was carried out on the bovine and porcine proteomes using NetPicoRNA, resulting in a list of 'potential' candidates. Table 5.2 shows how each data set was refined from the total list of 'potential' candidates to a list of characterised candidates and a list of characterised candidates with a cleavage score of 0.8 or greater. This second list was created for gene clustering analysis using the PANTHER and ToppFun systems.

Data set	Total candidates	Number of characterised candidates	Number of candidates with a cleavage score > 0.8
Bovine 3C candidates	969	426	126
Porcine 3C candidates	942	127	36

Table 5.2.: Table showing refinement of data sets after removal of uncharacterised proteins and implementation of cut off points

Predicted and characterised 3C protease targets were rated by their cleavage and surface scores. The ten highest rated genes are shown in Tables 5.3 and 5.4 while the full data sets are shown in Appendix A.6. None of the top ten candidates are found in both species and, indeed, there is very little evidence in the whole characterised data set of predicted targets which are common to the two species. The two 3C targets selected from the literature for the TALEN work, eIF4A1 and IKBKG, are also noticeably absent from both lists. Possible reasons for this are discussed in Section 5.4.

Gene clustering using PANTHER was carried out using both the full list of characterised candidates and the subset of candidates with a cleavage score of 0.8 or greater in addition to the complete bovine and porcine proteomes (Figures 5.11 and 5.12). The data were visualised as pie charts to determine if a particular molecular function or biological process is targeted by the 3C protease. Only slight changes were observed between the data sets, such as the decreases in protein binding transcription factor activity and structural molecule activity in the bovine molecular function analysis and the respective increase and decrease in the stimulus response and biological regulation categories in the porcine biological process analysis. Given that each of the three pie charts in each analysis represents a decreasing number of genes going from left to right, it seems unlikely that these changes have any biological significance.

The ToppFun analysis of the bovine data (Appendix 7.1) only allowed the candidate genes to be grouped by Pathway, and not Biological Process or Molecular Function and did not include any pathways which are known to be linked with FMDV replication. The porcine data is presented in Appendices 7.2, 7.3 and 7.4. Appendix 7.2 shows the potential candidates grouped by biological process. Interestingly, TRIM26 and TRIM31 appear twice in the data under the processes 'negative regulation of viral entry into host cell' and 'regulation of viral release from host cell', suggest that they could indeed be targets of the FMDV 3C protease. This result is discussed further in Section 5.4.1.

Protein description	Function	Uniprot identifier	P1 amino acid of cleavage site	Cleavage score of cleavage site	Surface score of cleavage site	Sequence of cleavage site	Likelihood of cleavage site
COL8A1 protein	Collagen chain	A7E303	Q196	0.956	0.76	IPGPQGPPG	Potential
DIS3-like exonuclease 2	3'-5'-exoribonuclease which mediates polyuridylated RNA degradation	E1B7R0	Q357	0.949	0.799	ECLPQGLPW	Potential
Collagen alpha-1(I) chain	Collagen chain	P02453	Q618	0.946	0.765	EAGAQGPPG	Potential
Chordin-like 2	Negative regulator of cartilage formation	Q2KJ20	Q53	0.943	0.766	YLEPQGLMY	Potential
LASS1 protein	Endoplasmic reticulum protein which synthesises ceramide	A4FUG2	E385	0.94	0.764	YALSEGPTG	Potential
1-aminocyclopropane-1-carboxylate synthase	Catalyses the deamination of L-vinylglycine	Q0V8M2	Q554	0.939	0.757	IQEPQGPHR	Potential
PAQR8 protein	Progesterone-binding membrane receptor	A6QLU0	E100	0.938	0.788	FVEAEGLPW	Potential
DNA-(apurinic or apyrimidinic site) lyase 2	Endodeoxyribonuclease involved in the DNA base excision repair pathway	F1MSK4	Q495	0.935	0.77	CARPQGPPT	Potential
Neurogenic differentiation factor	Transcriptional activator involved in neural differentiation	F1N2Z3	Q17	0.934	0.754	EPQPQGPPS	Potential
Pyruvate dehydrogenase (Lipoamide) alpha 2	Part of the pyruvate dehydrogenase complex which converts pyruvate to acetyl-CoA and CO_2	Q2T9Y3	E49	0.93	0.791	YRLEEGPPV	Potential

Table 5.3.: Top ten predicted targets in the bovine proteome for cleavage by FMDV 3C protease with predicted cleavage sites.

Protein description	Function	Uniprot identifier	P1 amino acid of cleavage site	Cleavage score of cleavage site	Surface score of cleavage site	Sequence of cleavage site	Likelihood of cleavage site
Oxysterol-binding protein (Fragment)	Involved in cholesterol metabolism	I3LKX6	E305	0.856	0.764	EPLKETTYT	Potential
Neuroendocrine convertase 2	Processes precursors of proteins such as hormones	F1SBI5	E64	0.845	0.787	LPFAEGLYH	Potential
Aquaporin 6	Transmembrane water channel	B2MUK1	E263	0.844	0.757	EPQKEGSQA	Potential
Sodium channel protein	Mediates sodium permeability of membranes	F1RSK2	E1764	0.842	0.782	APEKEGLIA	Potential
Oxysterol-binding protein	Involved in cholesterol metabolism	F1RY89	E60	0.841	0.798	PPRDEGPPT	Potential
Lysoplasmalogenase	Catalyses lysoplasmalogen degradation	F1RMN2	E6	0.839	0.777	DPGKEGLPR	Potential
Complement component MASP3	Part of the complement system	D5L7X4	E580	0.836	0.828	RPEPEGPAP	Potential
Sodium channel protein	Mediates sodium permeability of membranes	F1SJS7	E665	0.834	0.768	DPFAELTIT	Potential
Integrin beta-1-binding protein 2	May be involved in muscle cell maturation	Q462R2	E76	0.832	0.819	APQPEGPAT	Potential
Galectin	Binds lactose and other sugars	I3LGZ8	E181	0.83	0.777	LPSMEGSPT	Potential

Table 5.4.: Top ten predicted targets in the porcine proteome for cleavage by FMDV 3C protease with predicted cleavage sites.

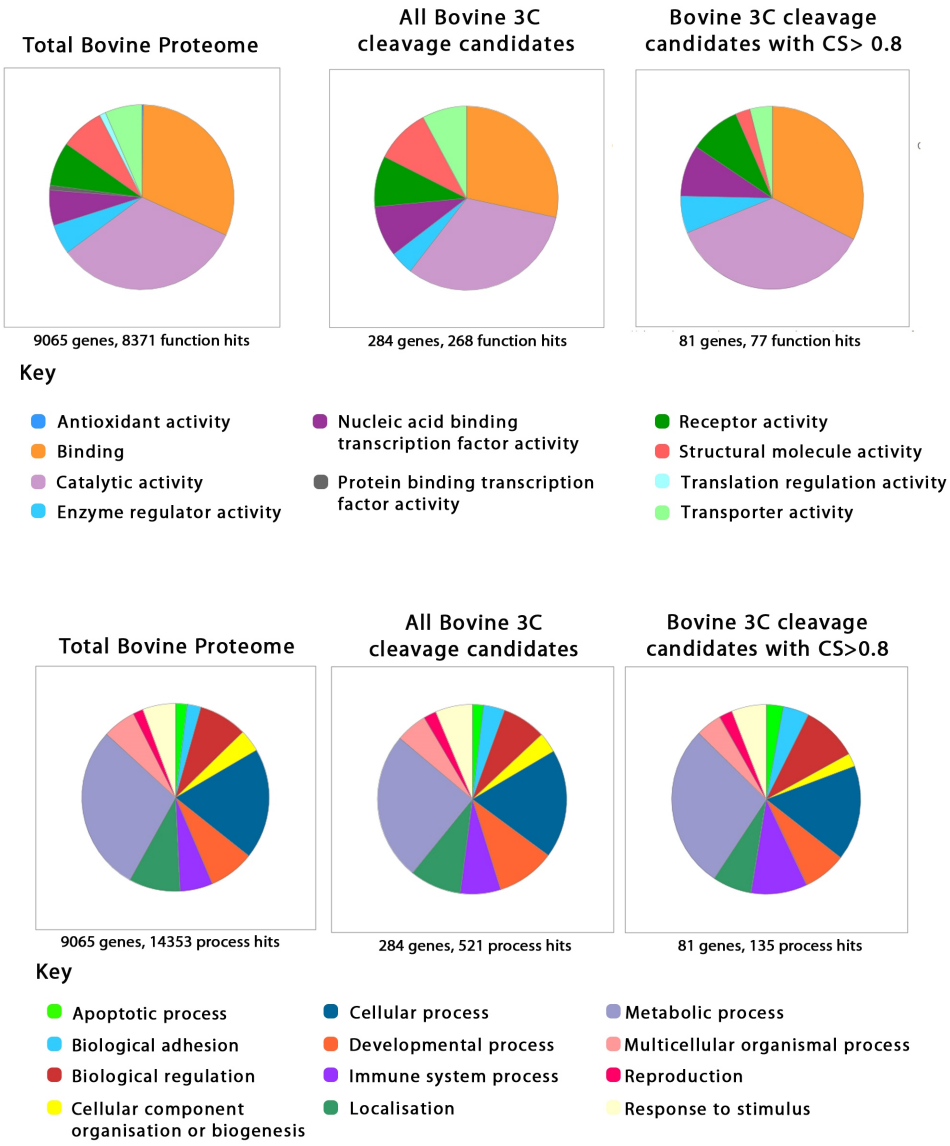


Figure 5.11.: PantherDB pie charts showing distribution of molecular functions (top) or biological processes (bottom) across bovine candidates for cleavage by the 3C protease. Analysis of the total bovine proteome is provided for comparison and keys are shown below their respective pie charts. The number of genes recognised by Panther DB and the total number of hits is given below each pie chart.

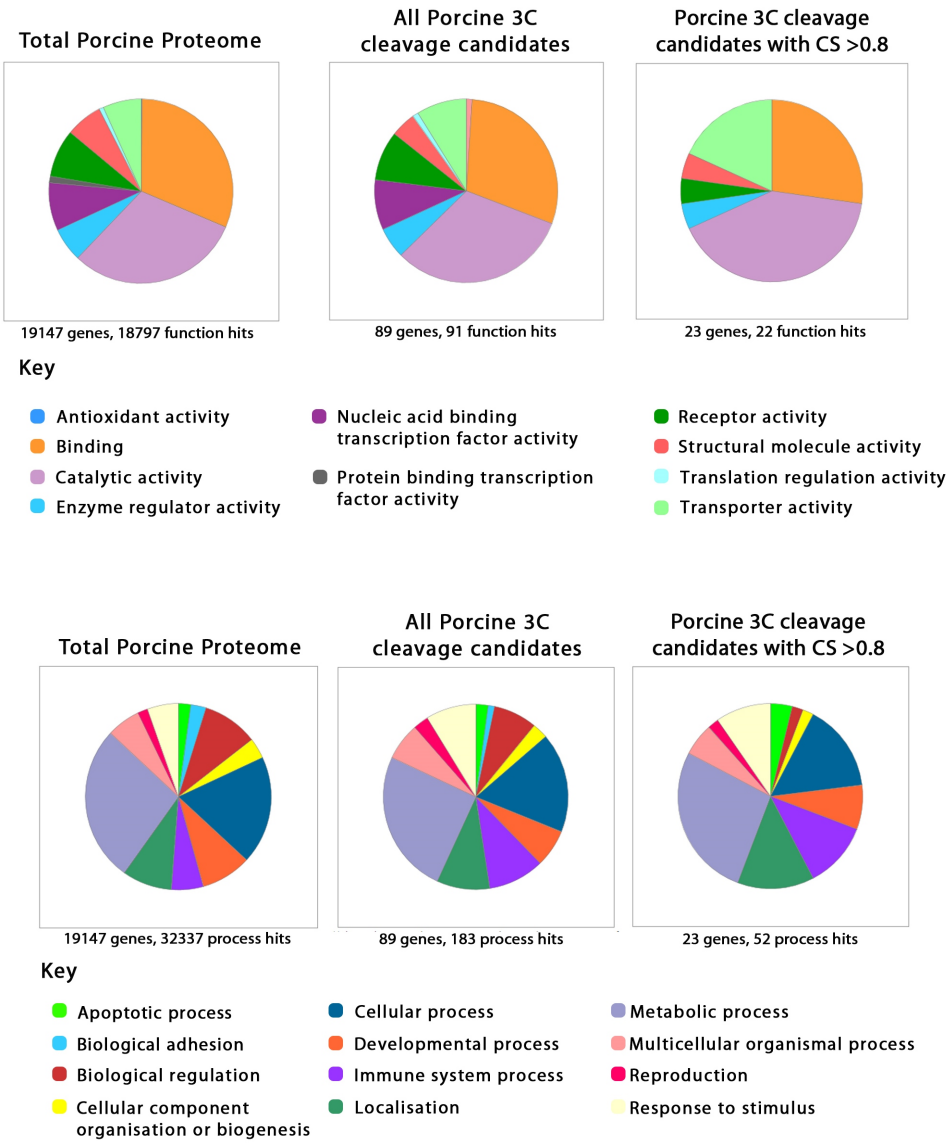


Figure 5.12.: PantherDB pie charts showing distribution of molecular functions (top) or biological processes (bottom) across porcine candidates for cleavage by the 3C protease. Analysis of the total porcine proteome is provided for comparison and keys are shown below their respective pie charts. The number of genes recognised by Panther DB and the total number of hits is given below each pie chart.

5.4. Discussion

This discussion section will focus on the bioinformatics work and the potential next steps for this work. A discussion of the validation issues with TALENs can be found in Chapter 6.

5.4.1. *In silico* cleavage site prediction and potential improvements

Bioinformatic analysis using NetPicoRNA was used to identify potential targets of the FMDV 3C protease in the bovine and porcine proteomes, with a view to producing a list of future candidates for genome editing. NetPicoRNA employs a neural network algorithm which was trained using known human targets of the FMDV 3C protease to predict potential cleavage sites.

Subsequent analysis of the list of targets using ToppFun indicated that the E3 ubiquitin ligases TRIM26 and TRIM31 may interact with the FMDV 3C protease in their roles controlling viral entry and viral release from a host cell. TRIM31 and TRIM26 are key components in innate immunity (Kawai and Akira, 2011), a role which renders them likely targets for degradation by FMDV. According to the NetPicoRNA analysis, TRIM26 could be cleaved by 3C at E385, while TRIM31 is potentially cleaved at E467. Both cleavage sites are located within the SPRY domains of each protein. SPRY domains are protein-interacting modules (as reviewed by D'Cruz et al. (2013)) so it could be conceived that cleavage and loss of SPRY from TRIM26 and TRIM31 would disrupt signalling cascades involved in the innate immune response. It would be of great use to determine whether these proteins are indeed targeted by the 3C protease by carrying out *in vitro* digests of TRIM protein with the 3C protease.

However, as stated in Section 5.3.3, none of the TALEN targets which were identified in the literature were present in the data set produced by NetPicoRNA. This could be due to either the algorithm itself or the relative lack of annotation in the bovine and porcine proteomes in comparison to others, such as human, meaning that relevant targets are still uncharacterised within the Uniprot databases and were therefore removed from the candidate list during data cleanup.

As the NetPicoRNA program was trained on human sequences, it is possible that many of the features that were used to create the original algorithms are not completely applicable to sequences from livestock species. The validation of the aphthovirus 3C algorithm also does

not appear to be as detailed as that of the other algorithms described in Blom et al. (1996) and there is no indication that the algorithms have been updated to include more recent data from the published literature.

Given the large number of proteins which appear to be targeted by the L protease (discussed further in Section 5.4.2), it would be of great use if a similar prediction algorithm could be produced for this enzyme. Previous work on the crystal structure of L protease has identified preferences for a leucine residue at position P2 of the cleavage site and for basic amino acids (i.e. arginine, lysine or histidine) at positions P'1 or P'2 (Guarne et al., 1998, Seipelt et al., 1999) as well as an inability to contain large amino acid side chains, such as phenylalanine, in its active site (Mayer et al., 2008). Further *in silico* analysis has suggested that the presence of a RKAR/RRLR motif may also indicate a potential L protease cleavage site (Pineiro et al., 2012). Potential consensus cleavage sites could also be tested *in vitro* by inserting the sequence of interest into a pro-GFP plasmid and co-transfection into cells with a plasmid containing the L or 3C protease (Callahan et al., 2010). The expression of GFP in the transfected cells would give a clear readout of the efficiency of cleavage.

Potential candidates could also be screened to determine the position of the predicted cleavage site within the protein's secondary structures, using a tool such as PSIPRED (McGuffin et al., 2000). Sequences situated on exposed loops are more easily accessed and cleaved by a protease than those in an α-helix or β-strand. This could also be used to help determine whether the predicted cleavage event removes a specific domain, as is the case in eIF4G1 cleavage by the L protease, or simply destroys the protein, as with 3C's cleavage of IKBKG. It will also be important to determine that the candidate protein is expressed in epithelial cells.

5.4.2. Identification of other potential genome editing candidates from the literature

Although only three targets were selected from the literature for genome editing in this work, the literature shows that there are, in theory, many different proteins which could be edited to become protease-resistant and therefore increase host resistance to FMDV infection. This includes proteins involved in transcriptional regulation, such as Histone H3 (Falk et al., 1990), and those in the autophagy pathway, as reviewed by Klein and Jackson (2011).

Analysis of FMDV IRES activity in a number of cell lines has shown that the presence

of the L protease can cause an increase in IRES activity, suggesting that the protease may have other targets than eIF4G1 that, upon cleavage, are subverted for use in virus replication (Roberts et al., 1998). Amongst these is the Gemin5 protein, which, when intact, binds to the FMDV IRES and exerts an inhibitory effect on FMDV translation. Cleavage by the L protease removes this inhibition (Pineiro et al., 2012). Another target which can increase IRES activity is the nuclear RNA binding protein Sam68 which is cleaved by the 3C protease (Lawrence et al., 2012). Cleavage of Sam68 causes the protein to move from the nucleus to the cytoplasm, where it interacts with the IRES. Knockdown of the protein causes a reduction in IRES activity and in viral titre, suggesting that the cleaved product facilitates FMDV replication (Lawrence et al., 2012).

Other targets may potentially be identified by looking at the targets of other picornaviruses however, this does not guarantee that the protein is also targeted by FMDV. An example of this are nucleoporin proteins, which are cleaved by rhinovirus proteases, causing inhibition of nuclear import and subsequent mislocalisation of proteins (Walker et al., 2013). However, analysis of FMDV-infected cells have shown only minor degradation of nucleoporin proteins (Lawrence et al., 2012).

The TALENs described in this chapter were designed against the ovine and porcine sequences however, there would be great value in seeking to modify the same proteins, IKBKG, eIF4A1 and eIF4G1, in cattle, as they appear to play a significant role in virus transmission in mixed populations (Bravo de Rueda et al., 2014a,b). As it is unlikely that FMDV-resistant animals from all susceptible livestock species would be developed at the same time, priority should be placed on those species which are both known to transmit the virus easily and can maintain a carrier status.

5.4.2.1. FMDV protease targeting of pathways involved in the innate immune response

IKBKG is just one of many proteins involved in signalling pathways of the innate immune response which have also been identified as targets of either FMDV or picornaviral proteases. The L protease, in particular, appears to play a key role in the inhibition of the IFNα/β immune response, suggesting that there are multiple targets affected by both the proteolytic and deubiquitinising properties of the enzyme (Chinsangaram et al., 1999, de Los Santos et al.,

2006). These include nuclear and cytoplasmic NF-κB protein (de Los Santos et al., 2007) as well as the transcription factors IRF3 and IRF7 (Wang et al., 2010), resulting in a considerable decrease in the expression of both interferon genes and genes containing ISREs, such as CCL5 (Wang et al., 2011a).

Several studies have shown that replacement of these degraded proteins by ectopic overexpression or use of constitutively active mutants can improve the innate immune response to FMDV infection (Xu et al., 2014, Ramirez-Carvajal et al., 2014, Shi et al., 2015), suggesting that preventing the cleavage of some of the affected pathway components could have a significant effect on FMDV pathology in livestock animals. Identifying potential targets from these pathways could prove to be an interesting avenue for future work.

5.4.3. An alternative method for engineering FMDV resistance

The transgenic chickens described in Lyall et al. (2011) could provide an alternative method for suppressing FMDV infection by using a short hairpin RNA (shRNA) targeting part of the FMDV genome to inhibit viral replication. shRNAs targeting the FMDV IRES or the VP1 region can reduce FMDV replication in cells and young mice (Wang et al., 2013b, Chang et al., 2014) and could be incorporated into a expression construct and introgressed into the host genome. However, as recent results in zebrafish have demonstrated, any candidate shRNA would have to first be checked to ensure that it does not exert any phenotypic effects on the host animal (Kok et al., 2015).

6. Conclusions and discussion

6.1. Genome editor validation issues

Table 6.1 gives an overview of the genome editors used in this project. Because it was not possible to accurately quantify the *in vitro* activity of the Rag1 and MSTN TALENs, due to oversaturation of the uncut bands on the gel image, an estimation of the level of activity is given.

The absence of detectable activity during the validation of genome editors has been a recurring issue in this project. Generally, multiple TALEN pairs or sgRNAs are designed against a single target with the expectation that at least one of the designed editors will exhibit an acceptable level of activity. 25 TALEN pairs and eight sgRNAs were used during the course of this work and only two TALEN pairs, Rag1 C (Figure 3.14) and MSTN (Figure 4.4), were shown to have activity *in vitro*. Every target had three TALEN pairs or four sgRNAs designed against it, as we have found this to be enough to produce at least one editor with an acceptable cutting efficiency in our lab. However, other labs have reported designing a larger number of editors against a single target in order to find a suitable editor to work with. It is therefore probable that the design and validation of more editors, in combination with the use of new design tools which have been developed (described in greater detail below), would have resulted in the production of a successful editor.

However, the MSTN sgRNAs did produce an edited animal, despite not showing any *in vitro* activity and repeat analysis of both the Rag1 C and MSTN TALENs failed to detect any activity, suggesting an issue with the validation process rather than the design process. As Rag1 C and MSTN TALEN pairs were the first to be constructed and validated for the project, their reassessment (and resulting negative results) came at the same time that the other genome editors were been built. The fact that the loss of activity appeared at the same time

Application	Gene	Species	Genome editor system used	Number of editors designed and tested	*In vitro* activity seen in CelI/T7 analysis	*In vivo* efficiency (total number of embryos and animals tested shown in brackets)
Creating a vacant haematopoietic niche for blastocyst complementation	Runx1	Pig	TALEN	3	None detected	N/A
	Rag1	Pig	TALEN	3	Some activity seen in pair C**	No editing (8 embryos and 6 lambs)
		Sheep	TALEN	3	None detected	No editing (22 embryos)
		Sheep	CRISPR	4	None detected	No editing (1 lamb)
Improving livestock productivity by increasing muscle growth	MSTN	Cattle	TALEN	1*	High activity**	N/A
		Sheep	TALEN	1*	Some activity**	One edited lamb (12 lambs)
		Sheep	CRISPR	4	None detected	One edited lamb (4 lambs)
Editing host genes to improve host resistance to FMDV infection	IKBKG	Pig and sheep	TALEN	3 for each species	None detected	N/A
	eIF4A1	Pig and sheep	TALEN	3 for each species	None detected	N/A
	eIF4G1	Pig and sheep	TALEN	3 for each species	None detected	N/A

Table 6.1.: Summary of the genome editors used in this project, including efficiencies observed *in vitro* and *in vivo*. * indicates that the TALEN pair had been designed and validated outside of this project. ** indicates that the observed activity was subsequently lost in repeat analyses. 'N/A' indicates editors which were not tested *in vivo*.

for all editors used suggested that the same problem was affecting all the editors, rather than the lack of activity in each editor being due to an individual problem. Therefore, efforts were made to attempt to identify any problems in the construction and validation processes, rather than redesigning new editors which may have still been affected by general construction and validation issues.

All the destination vectors into which genome editors used in this project were cloned have been used extensively in our lab and have been shown to function well. In order to rule out subcloning errors as a causal factor for the loss of TALEN activity, the ELD/KKR vectors were also sequenced to ensure that the correct heterodimer was used for the correct TALEN (Figure 4.6) and no issues were found. Based on this, the vectors were ruled out as a potential problem, leaving the validation process itself as the only procedure common to all the editors. This was also suggested by the production of the CRISPR-edited lamb (Figure 4.13), in spite of the fact that the MSTN sgRNAs were never shown to be active *in vitro*.

Troubleshooting steps were undertaken to attempt to determine whether the validation assay was the reason for the lack of detectable editing activity *in vitro*. The majority of these steps were carried out with the Rag1 TALENs as the Rag1 C pair had previously demonstrated specific cleavage in OEFs (Figure 3.15). The negative results from the transfections carried out independently (Figure 3.15 (c)) and from the GFP-sorting of transfected cells indicate that neither the transfection protocol or the level of transfection efficiency were the cause for the loss of function. Switching from a CelI assay to a T7 assay did increase the sensitivity of the validation procedure, as the T7 endonuclease was capable of detecting SNPs within the transfected cells, but did not show that any of the genome editors were functional. These results have left the true nature of the problem unclear. This is compounded by the fact that the technology is still relatively new and few examples of non-functional editors have been published in the literature.

It has been shown that the presence of methylated cytosine residues on the target DNA can block TALEN binding and therefore prevented cleavage of the DNA by FokI (Kim et al., 2013a). It is thought to be unlikely that this is the cause for the lack of activity seen here as two target sites, Rag1 and MSTN, had been previously shown to be permissive to editing and all of the FMDV protease target sites are located in exons of genes which are expressed in the vast majority of cell types and would not be expected to be repressed by cytosine methylation.

It is possible that the editors which never demonstrated activity *in vitro* have very poor cutting efficiencies, below the level which can be detected by either the CelI or T7 assays. Redesigning the editors using newly developed tools with more stringent design parameters such as SAPTA, the Broad Institute sgRNA Designer or the CRISPR Design tool from MIT (Lin et al., 2014, Doench et al., 2014, Hsu et al., 2013) could produce editors with an increased activity. The MIT CRISPR Design tool in particular would be of great use for future work in the area of genome edited large animals, as it is capable of predicting potential off-target sites in the pig genome.

Although time constraints meant that redesigning the editors was not possible, the previously designed sgRNAs targeted to Rag1 and MSTN were analysed using the Broad Institute sgRNA Designer to determine their predicted efficiency. The results of this are shown in Table 6.2. The Designer scores sgRNA candidates based on numerous features, including the GC content, the identity of the variable nucleotide in the NGG PAM and the presence or absence of certain nucleotides at specific positions in the sgRNA (e.g. cytosine is favoured at position 20, while thymine is disfavoured). It should be noted that the design scoring is very stringent and that, while a score of >0.6 is considered to be a good indicator of a highly active sgRNA, some sgRNAs were demonstrated to be active *in vitro* received a score of <0.2. This means that the design score should be taken as an indication of an sgRNA's activity, rather than as a definitive assessment.

The low design scores of the Rag1 sgRNAs do, indeed, suggest that these sgRNAs have a low cutting efficiency and would benefit from being redesigned. Interestingly, the efficiencies

sgRNA name	sgRNA Designer score
Ovine Rag1 F1	0.034
Ovine Rag1 F2	0.068
Ovine Rag1 R1	0.036
Ovine Rag1 R2	0.074
MSTN F1	0.034
MSTN F2	0.455
MSTN R1	0.109
MSTN R2	0.499

Table 6.2.: Table showing Broad Institute sgRNA Designer scores assigned to the four sgRNAs designed against ovine Rag1 and the four sgRNAs designed against ovine MSTN. The designer scores sgRNAs on their predicted efficiency from 0-1

of two of the MSTN sgRNAs, F2 and R2, were scored relatively well. As the CRISPR-edited lamb was injected with a mixture of all four MSTN sgRNAs, it is not possible to determine exactly which of the sgRNAs produced the editing event, but the high scores assigned to the F2 and R2 sgRNAs suggest that they are more likely to have been the cause of the deletion seen in the animal.

An alternative solution for the non-functional TALENs could be to construct them using a different system, such as Platinum Gate (Sakuma et al., 2013), which has been shown to increase the efficiency of cleavage. The Platinum Gate system differs from the Golden Gate system used in this project as the Platinum Gate domains contain additional variations outwith the RVD amino acids, which more closely mimics the structure of a TALE in *Xanthomonas*. Other TALEN scaffold vectors are also available however a comparison of scaffolds in Xenopus embryos indicated that the ELD/KKR heterodimeric vectors used in this project has the highest efficiency and lowest toxicity (Nakajima and Yaoita, 2013).

A final solution to combat deficiencies in the validation procedure would be to use a different assay to validate the editors. *In vitro* CRISPR/Cas9 digests using Cas9 proteins complexed with sgRNA on amplified or plasmid DNA have been detailed in the literature (Cho et al., 2013). These digests have been successfully carried out in our lab and would indicate if any potential problems with the current assay system are cell- or transfection-based. It would also be possible to try a different cell-based method, such as the traffic light reporter described in Certo et al. (2011), which would circumvent any problems with the DNA extraction and subsequent endonuclease assays used in the current validation protocol as the cutting efficiency of an editor can be determined by flow cytometry.

6.2. Introduction of genome edited animals into the agricultural industry

The TALEN- and CRISPR-edited lambs described in Chapter 4 and in the literature (Proudfoot et al., 2014) demonstrate the potential benefits that genome editors can bring to the agricultural industry and highlights the need for legislators to begin the process of determining how this technology and the resulting animals will be regulated.

Uncertainties surrounding the regulation of GMOs for human or animal consumption in dif-

ferent countries present one of the biggest barriers in developing genetic modification technology for widespread commercial use (Lusser et al., 2012). The regulation of GMOs intended for consumption is highly variable between countries as governments can either choose to implement a regulatory framework which is process-based (i.e. that focuses on the methods used to produce the GMO and places less emphasis on the actual modification itself) or product-based, which looks at the effects of each new modification in an organism and is less concerned with the method of modification, provided that the method has been previously found to be safe (Jones, 2015). Different countries also define the term 'genetic modification' differently. The FDA in the United States defines GMOs as organisms which carry a transgene in their genome, a feature that may not be routinely found in genome-edited organisms, while the European Commission's definition includes any DNA alterations which have not been introduced through breeding or natural recombination of loci.

While it is clear that some form of regulation will be required for genome-edited animals used in agriculture, the current regulation covering GMOs is not suitable for either the technology of genome editors or the use of genetically modified livestock in general. Currently, the vast majority of GMO legislation relates to genetically modified crops rather than animals, with the issue of containment of pollen or seeds to prevent cross-contamination being one of the biggest issues. Unlike plants, which can spread seeds and pollen indiscriminately through the air, containment of genetically modified animals should be easily managed by the use of secure housing. Even if a modified animal were to escape into the surrounding environment and successfully breed, it is unlikely that the animal would carry a modification which would be considered detrimental in the wild population. This is in contrast to the concerns about genetic drift of traits such as herbicide resistance from genetically modified crops into other plant species.

An exception to this would be animals which contain an integrated CRISPR/Cas9 cassette in their genome as part of multigenic chain reaction (MCR) experiments Gantz and Bier (2015). If the integrated CRISPR/Cas9 cassette entered a wild population and spread, it is likely that most or all animals in that population would carry mutations in both alleles of the target gene within a few generations. Therefore, any animals involved in MCR experiments would require a highly secure level of containment to ensure that the risk of them escaping is as minimal as possible. However, at this moment in time, it seems unlikely that MCR will be used to produce

edited livestock species for use in agriculture as the presence of the integrated CRISPR/Cas9 cassette in an animal's genome would potentially pose too great a risk in the eyes of regulators.

New regulatory frameworks would also need to take into account the new possibilities presented by genome editors compared to those produced using traditional genetic modification methods, such as recombinant DNA or transgene integration using viral vectors, the current focus of GMO regulation. The vast improvements in efficiency afforded by genome editors mean that marker genes, such as those encoding antibiotic resistance, are no longer required. This is in addition to the fact that genome editors allow far more subtle modifications to be produced, such as the insertion or removal of a few nucleotides in an endogenous gene, potentially reflecting mutations which could naturally occur in the genome. Given the different definitions of 'genetic modification' used by different regulatory bodies, it is probable that, as with current GMOs, regulation will vary internationally.

Another key issue for regulators to consider is the determination of off-target events caused by a genome editor. Whole-genome sequencing is now at a stage where subtle differences, and therefore potential off-target events, between isogenic cell lines can be easily detected (Veres et al., 2014, Smith et al., 2014, Yang et al., 2014, Suzuki et al., 2014). However, it is extremely difficult to source a suitable isogenic control for large animals such as pigs or sheep. As such, it is almost impossible to accurately differentiate between mutations which have been induced by a genome editor and those which have spontaneously arisen in an embryo or animal. This will be improved by the development of assays which are able to clearly signal the presence of off-target events, such as the assay recently published by Wang et al. (2015). Their method uses an integrase-defective lentiviral vector to physically mark loci in the genome where a DSB has occurred, allowing integration of the vector into the genomic DNA. Although the current sensitivity of the assay remains an issue at present, it has the potential to clearly delineate mutations in the genome which are caused by off-target activity and could complement whole-genome sequencing to great effect.

Although there is currently no regulation in place which specifically mentions the use of genome editors or of genetically modified livestock in agriculture, steps have been taken to address this. At present, few GMOs are authorised to be grown in the EU and the area has some of the most rigid GMO regulations globally. This makes it difficult to foresee how genome edited livestock would be regulated, however, some progress is being made. The European

Food Safety Authority has undertaken assessments of the safety and environmental impacts of genetically modified animals and has made recommendations to the European Commission for the implementation of a future regulatory framework. However, there has been little mention of the use of genome editors (Jones, 2015) and it is currently unclear when guidance on this will be issued.

Regulatory decisions in the United States give conflicting views about how the production and use of genome-edited animals may be controlled. The AquAdvantage salmon (Du et al., 1992), produced by insertion of a transgene, has long been expected to become the first genetically modified animal to be approved by the FDA for human consumption. Progress has, however, stalled numerous times since the salmon was first submitted for approval in 1996. The FDA published a finding of 'No Significant Impact' as the result of an environmental risk assessment on the salmon in December 2012 (Maxmen, 2012), but has not given final approval or commented further since then. In stark contrast to this, the US Department of Agriculture has recently waived the need for some genome-edited crops containing targeted deletions to be subject to regulation (Camacho et al., 2014), citing the lack of inserted genetic material as the reason for exemption. Edited crops with genomic insertions or substitutions derived from an exogenous template would need to be assessed on an individual basis. The extremes of these two cases make it very difficult to predict the level of regulation which might be applied to genome-edited livestock.

A more relaxed form of regulation for genome edited organisms in comparison to current GMO legislation could prove beneficial for smaller biotechnology companies, who are currently hindered by the financial costs and time needed to successfully steer a newly developed GMO through the regulatory process (Lusser et al., 2012), and would encourage the use of genome editors in more niche species, such as goats, which are not considered to be profitable, but could be of great benefit to agriculture in the developing world. However, a complete absence of regulation, as appears to be the current case with genome edited crops, could be harmful to the public's perceptions of genome editing. The debate which has surrounded genetically modified crops since the mid 1990s has made this area very emotionally charged and an 'all or nothing' approach to regulated genome editing will only serve to reinforce the public's suspicions of the technology.

Genome editors clearly have the potential to be a powerful tool in combating food insecurity,

one of the greatest problems facing society today. Increasing the productivity of livestock either directly, as with the myostatin-edited sheep, or more indirectly, by improving disease resistance, will help to meet an ever increasing global demand for meat and other livestock products (Gerber et al., 2013). However, it should be recognised that the increasing demand for meat is not only fuelled by an increasing global population, but also a rising number of newly wealthy, middle-class citizens who aspire to the meat-based Western diet. This adoption of the Western diet in an increasing number of developing countries is contributing a rise in noncommunicable diseases, such as cancer and cardiovascular disease (WHO, 2011) and the majority of countries with the highest obesity rates globally are found in the Middle East and the Pacific Islands (Morgen and Sorensen, 2014). In conclusion, while genome edited livestock can be used to increase the supply of livestock products, the consequences of this increased supply should also be considered.

Bibliography

T. Abe, Y. Hanazono, and Y. Nagao. A long-term follow-up study on the engraftment of human hematopoietic stem cells in sheep. *Exp. Anim.*, 63(4):475–481, 2014.

R. Acharya, E. Fry, D. Stuart, G. Fox, D. Rowlands, and F. Brown. The three-dimensional structure of foot-and-mouth disease virus at 2.9 A resolution. *Nature*, 337(6209):709–716, Feb. 1989.

L. Adamo, O. Naveiras, P. L. Wenzel, S. McKinney-Freeman, P. J. Mack, J. Gracia-Sancho, A. Suchy-Dicey, M. Yoshimoto, M. W. Lensch, M. C. Yoder, G. Garcia-Cardena, and G. Q. Daley. Biomechanical forces promote embryonic haematopoiesis. *Nature*, 459(7250):1131–1135, June 2009.

A. Agrawal and D. G. Schatz. RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell*, 89(1):43–53, Apr. 1997.

P. Ahnesorg, P. Smith, and S. P. Jackson. XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell*, 124(2):301–313, Jan. 2006.

K. Akopiants, R.-Z. Zhou, S. Mohapatra, K. Valerie, S. P. Lees-Miller, K.-J. Lee, D. J. Chen, P. Revy, J.-P. de Villartay, and L. F. Povirk. Requirement for XLF/Cernunnos in alignment-based gap filling by DNA polymerases lambda and mu for nonhomologous end joining in human whole-cell extracts. *Nucleic Acids Res.*, 37(12):4055–4062, July 2009.

S. Alexandersen, Z. Zhang, and A. I. Donaldson. Aspects of the persistence of foot-and-mouth disease virus in animals–the carrier problem. *Microbes Infect.*, 4(10):1099–1110, Aug. 2002.

S. Allais, H. Leveziel, N. Payet-Duprat, J. F. Hocquette, J. Lepetit, S. Rousset, C. Denoyelle, C. Bernard-Capel, L. Journaux, A. Bonnot, and G. Renand. The two mutations, Q204x and nt821, of the myostatin gene affect carcass and meat quality in young heterozygous bulls of French beef breeds. *J. Anim. Sci.*, 88(2):446–454, Feb. 2010.

D. L. Allen and T. G. Unterman. Regulation of myostatin expression and myoblast differen-

tiation by FoxO and SMAD transcription factors. *Am. J. Physiol., Cell Physiol.*, 292(1): C188–199, Jan. 2007.

A. Alvarez-Quilon, A. Serrano-Benitez, J. Ariel Lieberman, C. Quintero, D. Sanchez-Gutierrez, L. M. Escudero, and F. Cortes-Ledesma. ATM specifically mediates repair of double-strand breaks with blocked DNA ends. *Nature Communications*, 5, Feb. 2014.

H. Amthor, R. Macharia, R. Navarrete, M. Schuelke, S. C. Brown, A. Otto, T. Voit, F. Muntoni, G. Vrbova, T. Partridge, P. Zammit, L. Bunger, and K. Patel. Lack of myostatin results in excessive muscle growth but impaired force generation. *Proc. Natl. Acad. Sci. U.S.A.*, 104 (6):1835–1840, Feb. 2007.

I. Antony-Debre, D. Bluteau, R. Itzykson, V. Baccini, A. Renneville, F. Boehlen, M. Morabito, N. Droin, C. Deswarte, Y. Chang, G. Leverger, E. Solary, W. Vainchenker, R. Favier, and H. Raslova. MYH10 protein expression in platelets as a biomarker of RUNX1 and FLI1 alterations. *Blood*, June 2012.

M. A. Ariano, R. B. Armstrong, and V. R. Edgerton. Hindlimb muscle fiber populations of five mammals. *J. Histochem. Cytochem.*, 21(1):51–55, Jan. 1973.

S. Arnould, P. Chames, C. Perez, E. Lacroix, A. Duclert, J.-C. Epinat, F. Stricher, A.-S. Petit, A. Patin, S. Guillier, S. Rolland, J. Prieto, F. J. Blanco, J. Bravo, G. Montoya, L. Serrano, P. Duchateau, and F. Paques. Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.*, 355(3):443–458, Jan. 2006.

S. Arnould, C. Perez, J.-P. Cabaniols, J. Smith, A. Gouble, S. Grizot, J.-C. Epinat, A. Duclert, P. Duchateau, and F. Paques. Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.*, 371(1):49–65, Aug. 2007.

J. Ashworth, G. K. Taylor, J. J. Havranek, S. A. Quadri, B. L. Stoddard, and D. Baker. Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.*, 38(16):5601–5608, Sept. 2010.

H. Azuma, N. Paulk, A. Ranade, C. Dorrell, M. Al-Dhalimy, E. Ellis, S. Strom, M. A. Kay, M. Finegold, and M. Grompe. Robust expansion of human hepatocytes in Fah-/-/Rag2-/-/Il2rg-/- mice. *Nat. Biotechnol.*, 25(8):903–910, Aug. 2007.

J. A. Baan, T. Kocsis, A. Keller-Pinter, G. Mueller, E. Zador, L. Dux, and L. Mendler. The

compact mutation of myostatin causes a glycolytic shift in the phenotype of fast skeletal muscles. *J. Histochem. Cytochem.*, 61(12):889–900, Dec. 2013.

G. M. Bablanian and M. J. Grubman. Characterization of the foot-and-mouth disease virus 3c protease expressed in Escherichia coli. *Virology*, 197(1):320–327, Nov. 1993.

C. J. Bakkenist and M. B. Kastan. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, 421(6922):499–506, Jan. 2003.

R. Bakshi, M. Q. Hassan, J. Pratap, J. B. Lian, M. A. Montecino, A. J. van Wijnen, J. L. Stein, A. N. Imbalzano, and G. S. Stein. The human SWI/SNF complex associates with RUNX1 to control transcription of hematopoietic target genes. *J. Cell. Physiol.*, 225(2):569–576, Nov. 2010.

K. M. Baldwin and C. M. Tipton. Work and metabolic patterns of fast and slow twitch skeletal muscle contracting in situ. *Pflugers Arch.*, 334(4):345–356, 1972.

C. Baligand, H. Gilson, J. C. MÃ©nard, O. Schakman, C. Wary, J.-P. Thissen, and P. G. Carlier. Functional assessment of skeletal muscle in intact mice lacking myostatin by concurrent NMR imaging and spectroscopy. *Gene Ther.*, 17(3):328–337, Mar. 2010.

L. J. Barber, J. L. Youds, J. D. Ward, M. J. McIlwraith, N. J. O'Neil, M. I. R. Petalcorin, J. S. Martin, S. J. Collis, S. B. Cantor, M. Auclair, H. Tissenbaum, S. C. West, A. M. Rose, and S. J. Boulton. RTEL1 maintains genomic stability by suppressing homologous recombination. *Cell*, 135(2):261–271, Oct. 2008.

R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(5819):1709–1712, Mar. 2007.

M. J. Barrero and J. C. Izpisua Belmonte. iPS cells forgive but do not forget. *Nat. Cell Biol.*, 13(5):523–525, May 2011.

V. M. Bedell, Y. Wang, J. M. Campbell, T. L. Poshusta, C. G. Starker, R. G. Krug, 2nd, W. Tan, S. G. Penheiter, A. C. Ma, A. Y. H. Leung, S. C. Fahrenkrug, D. F. Carlson, D. F. Voytas, K. J. Clark, J. J. Essner, and S. C. Ekker. In vivo genome editing using a high-efficiency TALEN system. *Nature*, 491(7422):114–118, Nov. 2012.

I. Bedzhov, C. Y. Leung, M. Bialecka, and M. Zernicka-Goetz. In vitro culture of mouse blastocysts beyond the implantation stages. *Nat Protoc*, 9(12):2732–2739, Dec. 2014.

T. Bee, K. Liddiard, G. Swiers, S. R. B. Bickley, C. S. Vink, A. Jarratt, J. R. Hughes, A. Medv-

insky, and M. F. T. R. de Bruijn. Alternative Runx1 promoter usage in mouse developmental hematopoiesis. *Blood Cells, Molecules & Diseases*, 43(1):35–42, Aug. 2009.

R. H. S. Bellinge, D. A. Liberles, S. P. A. Iaschi, P. A. O'brien, and G. K. Tay. Myostatin and its implications on animal breeding: a review. *Anim. Genet.*, 36(1):1–6, Feb. 2005.

G. J. Belsham and C. J. Bostock. Studies on the infectivity of foot-and-mouth disease virus RNA using microinjection. *J. Gen. Virol.*, 69 (Pt 2):265–274, Feb. 1988.

G. J. Belsham and J. K. Brangwyn. A region of the 5' noncoding region of foot-and-mouth disease virus RNA directs efficient internal initiation of protein synthesis within cells: involvement with the role of L protease in translational control. *J. Virol.*, 64(11):5389–5395, Nov. 1990.

G. J. Belsham, G. M. McInerney, and N. Ross-Smith. Foot-and-mouth disease virus 3c protease induces cleavage of translation initiation factors eIF4a and eIF4g within infected cells. *J. Virol.*, 74(1):272–280, Jan. 2000.

K. L. Berkner. Development of adenovirus vectors for the expression of heterologous genes. *BioTechniques*, 6(7):616–629, Aug. 1988.

F. Bernardin-Fried, T. Kummalue, S. Leijen, M. I. Collector, K. Ravid, and A. D. Friedman. AML1/RUNX1 Increases During G1 to S Cell Cycle Progression Independent of Cytokine-dependent Phosphorylation and Induces Cyclin D3 Gene Expression. *Journal of Biological Chemistry*, 279(15):15678–15687, Apr. 2004.

S. Berryman, S. Clark, P. Monaghan, and T. Jackson. Early events in integrin alphavbeta6-mediated cell entry of foot-and-mouth disease virus. *J. Virol.*, 79(13):8519–8534, July 2005.

M. W. Bevan, R. B. Flavell, and M.-D. Chilton. A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature*, 304(5922):184–187, July 1983.

D. Bhaya, M. Davison, and R. Barrangou. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.*, 45:273–297, 2011.

M. Bibikova, D. Carroll, D. J. Segal, J. K. Trautman, J. Smith, Y. G. Kim, and S. Chandrasegaran. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.*, 21(1):289–297, Jan. 2001.

M. Bibikova, K. Beumer, J. K. Trautman, and D. Carroll. Enhancing gene targeting with

designed zinc finger nucleases. *Science*, 300(5620):764, May 2003. ISSN 1095-9203. doi: 10.1126/science.1079512.

J. R. Birtley, S. R. Knox, A. M. Jaulent, P. Brick, R. J. Leatherbarrow, and S. Curry. Crystal structure of foot-and-mouth disease virus 3c protease. New insights into catalytic mechanism and cleavage specificity. *J. Biol. Chem.*, 280(12):11520–11527, Mar. 2005.

J. Bitinaite, D. A. Wah, A. K. Aggarwal, and I. Schildkraut. FokI dimerization is required for DNA cleavage. *Proc. Natl. Acad. Sci. U.S.A.*, 95(18):10570–10575, Sept. 1998.

K. Blair, H. G. Leitch, W. Mansfield, C.-E. Dumeau, P. Humphreys, and A. G. Smith. Culture parameters for stable expansion, genetic modification and germline transmission of rat pluripotent stem cells. *Biol Open*, 1(1):58–65, Jan. 2012.

R. B. Blasco, E. Karaca, C. Ambrogio, T.-C. Cheong, E. Karayol, V. G. Minero, C. Voena, and R. Chiarle. Simple and Rapid In Vivo Generation of Chromosomal Rearrangements using CRISPR/Cas9 Technology. *Cell Rep*, 9(4):1219–1227, Nov. 2014.

N. Blom, J. Hansen, D. Blaas, and S. Brunak. Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Sci.*, 5(11):2203–2216, Nov. 1996.

J. Boch and U. Bonas. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol*, 48:419–436, 2010.

J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326(5959):1509–1512, Dec. 2009.

A. J. Bogdanove and D. F. Voytas. TAL effectors: customizable proteins for DNA targeting. *Science*, 333(6051):1843–1846, Sept. 2011.

J.-C. Boisset, W. van Cappellen, C. Andrieu-Soler, N. Galjart, E. Dzierzak, and C. Robin. In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium. *Nature*, 464(7285):116–120, Mar. 2010.

A. Bolotin, B. Quinquis, A. Sorokin, and S. D. Ehrlich. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading, Engl.)*, 151(Pt 8):2551–2561, Aug. 2005.

I. A. Boman, G. Klemetsdal, T. Blichfeldt, O. Nafstad, and D. I. VÃ¥ge. A frameshift mutation

in the coding region of the myostatin gene (MSTN) affects carcass conformation and fatness in Norwegian White Sheep (Ovis aries). *Anim. Genet.*, 40(4):418–422, Aug. 2009.

I. A. Boman, G. Klemetsdal, O. Nafstad, T. Blichfeldt, and D. I. VÃ¥ge. Selection based on progeny testing induces rapid changes in myostatin allele frequencies - a case study in sheep. *J. Anim. Breed. Genet.*, 128(1):52–55, Feb. 2011.

U. Bonas, R. E. Stall, and B. Staskawicz. Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria. *Mol. Gen. Genet.*, 218 (1):127–136, July 1989.

R. A. Bosselman, R. Y. Hsu, T. Boggs, S. Hu, J. Bruszewski, S. Ou, L. Kozar, F. Martin, C. Green, and F. Jacobsen. Germline transmission of exogenous genes in the chicken. *Science*, 243(4890):533–535, Jan. 1989.

S. J. Boulton and S. P. Jackson. Saccharomyces cerevisiae Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.*, 15(18):5093–5103, Sept. 1996.

A. P. Boyle, C. L. Araya, C. Brdlik, P. Cayting, C. Cheng, Y. Cheng, K. Gardner, L. W. Hillier, J. Janette, L. Jiang, D. Kasper, T. Kawli, P. Kheradpour, A. Kundaje, J. J. Li, L. Ma, W. Niu, E. J. Rehm, J. Rozowsky, M. Slattery, R. Spokony, R. Terrell, D. Vafeados, D. Wang, P. Weisdepp, Y.-C. Wu, D. Xie, K.-K. Yan, E. A. Feingold, P. J. Good, M. J. Pazin, H. Huang, P. J. Bickel, S. E. Brenner, V. Reinke, R. H. Waterston, M. Gerstein, K. P. White, M. Kellis, and M. Snyder. Comparative analysis of regulatory information and circuits across distant species. *Nature*, 512(7515):453–456, Aug. 2014.

A. Bradley, M. Evans, M. H. Kaufman, and E. Robertson. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309(5965):255–256, May 1984.

M. Braga, S. Pervin, K. Norris, S. Bhasin, and R. Singh. Inhibition of in vitro and in vivo brown fat differentiation program by myostatin. *Obesity (Silver Spring)*, 21(6):1180–1188, June 2013.

J. Bravo, Z. Li, N. A. Speck, and A. J. Warren. The leukemia-associated AML1 (Runx1)–CBF beta complex functions as a DNA-induced molecular clamp. *Nat. Struct. Biol.*, 8(4): 371–378, Apr. 2001.

C. Bravo de Rueda, M. C. M. de Jong, P. L. EblÃ©, and A. Dekker. Estimation of the transmission of foot-and-mouth disease virus from infected sheep to cattle. *Vet. Res.*, 45:58, 2014a.

C. Bravo de Rueda, A. Dekker, P. L. Eble, and M. C. M. de Jong. Vaccination of cattle only is sufficient to stop FMDV transmission in mixed populations of sheep and cattle. *Epidemiol. Infect.*, pages 1–8, Dec. 2014b.

A. L. Bredemeyer, G. G. Sharma, C.-Y. Huang, B. A. Helmink, L. M. Walker, K. C. Khor, B. Nuskey, K. E. Sullivan, T. K. Pandita, C. H. Bassing, and B. P. Sleckman. ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature*, 442(7101):466–470, July 2006.

G. Brem, H. Tenhumberg, and H. KrÃ€usslich. Chimerism in cattle through microsurgical aggregation of morulae. *Theriogenology*, 22(5):609–613, Nov. 1984.

D. Brenin, J. Look, M. Bader, N. HÃŒbner, G. Levan, and P. Iannaccone. Rat embryonic stem cells: a progress report. *Transplant. Proc.*, 29(3):1761–1765, May 1997.

B. Brophy, G. Smolenski, T. Wheeler, D. Wells, P. L'Huillier, and G. Laible. Cloned transgenic cattle produce milk with higher levels of beta-casein and kappa-casein. *Nat. Biotechnol.*, 21 (2):157–162, Feb. 2003.

S. J. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. H. Slijkhuis, A. P. L. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. van der Oost. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891):960–964, Aug. 2008.

R. S. Brown, C. Sander, and P. Argos. The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Letters*, 186(2):271–274, July 1985.

M. Buehr, S. Meek, K. Blair, J. Yang, J. Ure, J. Silva, R. McLay, J. Hall, Q.-L. Ying, and A. Smith. Capture of authentic embryonic stem cells from rat blastocysts. *Cell*, 135(7): 1287–1298, Dec. 2008.

J. M. Butler, D. J. Nolan, E. L. Vertes, B. Varnum-Finney, H. Kobayashi, A. T. Hooper, M. Seandel, K. Shido, I. A. White, M. Kobayashi, L. Witte, C. May, C. Shawber, Y. Kimura, J. Kitajewski, Z. Rosenwaks, I. D. Bernstein, and S. Rafii. Endothelial cells are essential for the self-renewal and repopulation of Notch-dependent hematopoietic stem cells. *Cell Stem Cell*, 6(3):251–264, Mar. 2010.

X. Cai, J. J. Gaudet, J. K. Mangan, M. J. Chen, M. E. De Obaldia, Z. Oo, P. Ernst, and N. A. Speck. Runx1 loss minimally impacts long-term hematopoietic stem cells. *PLoS ONE*, 6 (12):e28430, 2011.

Z. Cai, M. de Bruijn, X. Ma, B. Dortland, T. Luteijn, R. J. Downing, and E. Dzierzak. Hap-

loinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity*, 13(4):423–431, Oct. 2000.

B. P. Callahan, M. J. Stanger, and M. Belfort. Protease Activation of Split Green Fluorescent Protein. *ChemBioChem*, 11(16):2259–2263, Nov. 2010.

A. Camacho, A. Van Deynze, C. Chi-Ham, and A. B. Bennett. Genetically engineered crops that fly under the US regulatory radar. *Nat. Biotechnol.*, 32(11):1087–1091, Nov. 2014. ISSN 1546-1696.

J. Canon and U. Banerjee. Runt and Lozenge function in Drosophila development. *Semin. Cell Dev. Biol.*, 11(5):327–336, Oct. 2000.

X. Cao, I. E. Bergmann, R. FÃŒllkrug, and E. Beck. Functional analysis of the two alternative translation initiation sites of foot-and-mouth disease virus. *J. Virol.*, 69(1):560–563, Jan. 1995.

D. F. Carlson, W. Tan, S. G. Lillico, D. Stverakova, C. Proudfoot, M. Christian, D. F. Voytas, C. R. Long, C. B. A. Whitelaw, and S. C. Fahrenkrug. Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl. Acad. Sci. U.S.A.*, 109(43):17382–17387, Oct. 2012.

E. C. Carrillo, C. Giachetti, and R. H. Campos. Effect of lysosomotropic agents on the foot-and-mouth disease virus replication. *Virology*, 135(2):542–545, June 1984.

E. C. Carrillo, C. Giachetti, and R. Campos. Early steps in FMDV replication: further analysis on the effects of chloroquine. *Virology*, 147(1):118–125, Nov. 1985.

A. Carver, G. Wright, D. Cottom, J. Cooper, M. Dalrymple, S. Temperley, M. Udell, D. Reeves, J. Percy, and A. Scott. Expression of human alpha 1 antitrypsin in transgenic sheep. *Cytotechnology*, 9(1-3):77–84, 1992.

J. N. Cash, C. A. Rejon, A. C. McPherron, D. J. Bernard, and T. B. Thompson. The structure of myostatin:follistatin 288: insights into receptor utilization and heparin binding. *EMBO J.*, 28(17):2662–2676, Sept. 2009.

J. N. Cash, E. B. Angerman, C. Kattamuri, K. Nolan, H. Zhao, Y. Sidis, H. T. Keutmann, and T. B. Thompson. Structure of myostatinÂ·follistatin-like 3: N-terminal domains of follistatin-type molecules exhibit alternate modes of binding. *J. Biol. Chem.*, 287(2):1043–1053, Jan. 2012.

T. Cathomen and J. K. Joung. Zinc-finger Nucleases: The Next Generation Emerges. *Molecular Therapy*, 16(7):1200–1207, July 2008.

A. Celeste, S. Petersen, P. J. Romanienko, O. Fernandez-Capetillo, H. T. Chen, O. A. Sedelnikova, B. Reina-San-Martin, V. Coppola, E. Meffre, M. J. Difilippantonio, C. Redon, D. R. Pilch, A. Olaru, M. Eckhaus, R. D. Camerini-Otero, L. Tessarollo, F. Livak, K. Manova, W. M. Bonner, M. C. Nussenzweig, and A. Nussenzweig. Genomic instability in mice lacking histone H2ax. *Science*, 296(5569):922–927, May 2002.

R. Cencic, H. Miura, A. Malina, F. Robert, S. Ethier, T. M. Schmeing, J. Dostie, and J. Pelletier. Protospacer Adjacent Motif (PAM)-Distal Sequences Engage CRISPR Cas9 DNA Target Cleavage. *PLoS ONE*, 9(10):e109213, 2014.

T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, and D. F. Voytas. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, 39(12): 82, July 2011.

M. T. Certo, B. Y. Ryu, J. E. Annis, M. Garibov, J. Jarjour, D. J. Rawlings, and A. M. Scharenberg. Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods*, 8(8):671–676, Aug. 2011.

M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805, Feb. 1994.

G. A. Challen and M. A. Goodell. Runx1 isoforms show differential expression patterns during hematopoietic development but have similar functional effects in adult hematopoietic stem cells. *Exp. Hematol.*, 38(5):403–416, May 2010.

N. Chamond, J. Deforges, N. Ulryck, and B. Sargueil. 40s recruitment in the absence of eIF4g/4a by EMCV IRES refines the model for translation initiation on the archetype of Type II IRESs. *Nucleic Acids Res.*, 42(16):10373–10384, 2014.

A. W. Chan, K. Y. Chong, C. Martinovich, C. Simerly, and G. Schatten. Transgenic monkeys produced by retroviral gene transfer into mature oocytes. *Science*, 291(5502):309–312, Jan. 2001.

Y. Chang, Y. Dou, H. Bao, X. Luo, X. Liu, K. Mu, Z. Liu, X. Liu, and X. Cai. Multiple microRNAs targeted to internal ribosome entry site against foot-and-mouth disease virus infection in vitro and in vivo. *Virol. J.*, 11:1, 2014.

P. Chapdelaine, C. Pichavant, J. Rousseau, F. Paques, and J. P. Tremblay. Meganucleases can restore the reading frame of a mutated dystrophin. *Gene Ther.*, 17(7):846–858, July 2010.

C. Chappell, L. A. Hanakahi, F. Karimi-Busheri, M. Weinfeld, and S. C. West. Involvement of human polynucleotide kinase in double-strand break repair by non-homologous end joining. *EMBO J.*, 21(11):2827–2832, June 2002.

P. Charbord, C. Pouget, H. Binder, F. Dumont, G. Stik, P. Levy, F. Allain, C. Marchal, J. Richter, B. Uzan, F. Pflumio, F. Letourneur, H. Wirth, E. Dzierzak, D. Traver, T. Jaffredo, and C. Durand. A systems biology approach for defining the molecular framework of the hematopoietic stem cell niche. *Cell Stem Cell*, 15(3):376–391, Sept. 2014.

A. J. Chase and B. L. Semler. Viral subversion of host functions for picornavirus translation and RNA replication. *Future Virol*, 7(2):179–191, Feb. 2012.

N. H. Chehab, A. Malikzay, M. Appel, and T. D. Halazonetis. Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53. *Genes Dev.*, 14(3):278–288, Feb. 2000.

C.-L. Chen, D. C. Broom, Y. Liu, J. C. de Nooij, Z. Li, C. Cen, O. A. Samad, T. M. Jessell, C. J. Woolf, and Q. Ma. Runx1 determines nociceptive sensory neuron phenotype and is required for thermal and neuropathic pain. *Neuron*, 49(3):365–377, Feb. 2006.

J. Chen, R. Lansford, V. Stewart, F. Young, and F. W. Alt. RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proc. Natl. Acad. Sci. U.S.A.*, 90(10):4528–4532, May 1993.

J. Chen, E. E. Bardes, B. J. Aronow, and A. G. Jegga. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.*, 37(Web Server issue):W305–311, July 2009a. ISSN 1362-4962.

M. J. Chen, T. Yokomizo, B. M. Zeigler, E. Dzierzak, and N. A. Speck. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*, 457(7231): 887–891, Feb. 2009b.

Z. Cheng, P. Yi, X. Wang, Y. Chai, G. Feng, Y. Yang, X. Liang, Z. Zhu, W. Li, and G. Ou. Conditional targeted genome editing using somatically expressed TALENs in C. elegans. *Nat. Biotechnol.*, 31(10):934–937, Oct. 2013.

B. S. Chevalier, T. Kortemme, M. S. Chadsey, D. Baker, R. J. Monnat, and B. L. Stoddard. Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell*, 10(4): 895–905, Oct. 2002.

J. Chinsangaram, M. E. Piccone, and M. J. Grubman. Ability of foot-and-mouth disease virus

to form plaques in cell culture is associated with suppression of alpha/beta interferon. *J. Virol.*, 73(12):9891–9898, Dec. 1999.

S. W. Cho, S. Kim, J. M. Kim, and J.-S. Kim. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*, Jan. 2013.

S. W. Cho, S. Kim, Y. Kim, J. Kweon, H. S. Kim, S. Bae, and J.-S. Kim. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research*, 24(1):132–141, Jan. 2014.

A. Choulika, A. Perrin, B. Dujon, and J. F. Nicolas. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol. Cell. Biol.*, 15(4):1968–1973, Apr. 1995.

M. Christian, T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, and D. F. Voytas. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186(2):757–761, Oct. 2010.

A. Ciccia and S. J. Elledge. The DNA Damage Response: Making It Safe to Play with Knives. *Molecular Cell*, 40(2):179–204, Oct. 2010.

A. Cieslak, S. Le Noir, A. Trinquand, L. Lhermitte, D.-M. Franchini, P. Villarese, S. Gon, J. Bond, M. Simonin, L. Vanhille, C. Reimann, E. Verhoeyen, J. Larghero, E. Six, S. Spicuglia, I. Andre-Schmutz, A. Langerak, B. Nadel, E. Macintyre, D. Payet-Bornet, and V. Asnafi. RUNX1-dependent RAG1 deposition instigates human TCR-delta locus re-arrangement. *J. Exp. Med.*, Nov. 2014.

M. Ciubotaru, A. J. Trexler, L. N. Spiridon, M. D. Surleac, E. Rhoades, A. J. Petrescu, and D. G. Schatz. RAG and HMGB1 create a large bend in the 23rss in the V(D)J recombination synaptic complexes. *Nucleic Acids Res.*, 41(4):2437–2454, Feb. 2013.

M. Ciubotaru, M. D. Surleac, L. A. Metskas, P. Koo, E. Rhoades, A. J. Petrescu, and D. G. Schatz. The architecture of the 12rss in V(D)J recombination signal and synaptic complexes. *Nucleic Acids Res.*, Dec. 2014.

W. K. Clements, A. D. Kim, K. G. Ong, J. C. Moore, N. D. Lawson, and D. Traver. A somitic Wnt16/Notch pathway specifies haematopoietic stem cells. *Nature*, 474(7350):220–224, June 2011.

A. Clop, F. Marcq, H. Takeda, D. Pirottin, X. Tordoir, B. BibÃ©, J. Bouix, F. Caiment, J.-M. Elsen, F. Eychenne, C. Larzul, E. Laville, F. Meish, D. Milenkovic, J. Tobin, C. Charlier,

and M. Georges. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.*, 38(7):813–818, July 2006.

S. N. Cohen, A. C. Chang, and L. Hsu. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 69(8):2110–2114, Aug. 1972.

S. N. Cohen, A. C. Chang, H. W. Boyer, and R. B. Helling. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, 70(11):3240–3244, Nov. 1973.

M. J. Coldwell, L. Hashemzadeh-Bonehi, T. M. Hinton, S. J. Morley, and V. M. Pain. Expression of fragments of translation initiation factor eIF4gi reveals a nuclear localisation signal within the N-terminal apoptotic cleavage fragment N-FAG. *J. Cell. Sci.*, 117(Pt 12): 2545–2555, May 2004.

L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121):819–823, Feb. 2013.

B. Corneo, R. L. Wendland, L. Deriano, X. Cui, I. A. Klein, S.-Y. Wong, S. Arnal, A. J. Holub, G. R. Weller, B. A. Pancake, S. Shah, V. L. Brandt, K. Meek, and D. B. Roth. Rag mutations reveal robust alternative end joining. *Nature*, 449(7161):483–486, Sept. 2007.

D. Cortez, Y. Wang, J. Qin, and S. J. Elledge. Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science*, 286(5442):1162–1166, Nov. 1999.

D. Coverley, M. K. Kenny, M. Munn, W. D. Rupp, D. P. Lane, and R. D. Wood. Requirement for the replication protein SSB in human DNA excision repair. *Nature*, 349(6309):538–541, Feb. 1991.

D. Coverley, M. K. Kenny, D. P. Lane, and R. D. Wood. A role for the human single-stranded DNA binding protein HSSB/RPA in an early stage of nucleotide excision repair. *Nucleic Acids Res.*, 20(15):3873–3880, Aug. 1992.

T. J. Cradick, E. J. Fine, C. J. Antico, and G. Bao. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.*, 41(20):9584–9592, Nov. 2013.

A. J. Davis and D. J. Chen. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res*, 2(3):130–143, June 2013.

A. J. Davis, L. Chi, S. So, K.-J. Lee, E. Mori, K. Fattah, J. Yang, and D. J. Chen. BRCA1 modulates the autophosphorylation status of DNA-PKcs in S phase of the cell cycle. *Nucleic Acids Res.*, 42(18):11487–11501, Feb. 2015.

A. A. D'Cruz, J. J. Babon, R. S. Norton, N. A. Nicola, and S. E. Nicholson. Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity. *Protein Sci.*, 22 (1):1–10, Jan. 2013.

M. de Jager, J. van Noort, D. C. van Gent, C. Dekker, R. Kanaar, and C. Wyman. Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol. Cell*, 8(5):1129–1135, Nov. 2001.

T. de Los Santos, S. de Avila Botton, R. Weiblen, and M. J. Grubman. The leader proteinase of foot-and-mouth disease virus inhibits the induction of beta interferon mRNA and blocks the host innate immune response. *J. Virol.*, 80(4):1906–1914, Feb. 2006.

T. de Los Santos, F. Diaz-San Segundo, and M. J. Grubman. Degradation of nuclear factor kappa B during foot-and-mouth disease virus infection. *J. Virol.*, 81(23):12803–12815, Dec. 2007.

L. G. DeFazio, R. M. Stansel, J. D. Griffith, and G. Chu. Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J.*, 21(12):3192–3200, June 2002.

E. Deltcheva, K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel, and E. Charpentier. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340):602–607, Mar. 2011.

L. Deriano, T. H. Stracker, A. Baker, J. H. J. Petrini, and D. B. Roth. Roles for NBS1 in alternative nonhomologous end-joining of V(D)J recombination intermediates. *Mol. Cell*, 34(1):13–25, Apr. 2009.

M. A. Devaney, V. N. Vakharia, R. E. Lloyd, E. Ehrenfeld, and M. J. Grubman. Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap-binding protein complex. *J. Virol.*, 62(11):4407–4409, Nov. 1988.

R. H. Devlin, T. Y. Yesaki, C. A. Biagi, E. M. Donaldson, P. Swanson, and W.-K. Chan. Extraordinary salmon growth. *Nature*, 371(6494):209–210, Sept. 1994.

T. M. Dexter, T. D. Allen, and L. G. Lajtha. Conditions controlling the proliferation of haemopoietic stem cells in vitro. *J. Cell. Physiol.*, 91(3):335–344, June 1977.

M. J. Difilippantonio, C. J. McMahan, Q. M. Eastman, E. Spanopoulou, and D. G. Schatz. RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell*, 87(2):253–262, Oct. 1996.

J. G. Doench, E. Hartenian, D. B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B. L. Ebert, R. J. Xavier, and D. E. Root. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.*, Sept. 2014.

T. Doetschman, R. G. Gregg, N. Maeda, M. L. Hooper, D. W. Melton, S. Thompson, and O. Smithies. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, 330(6148):576–578, Dec. 1987.

Y. Doyon, T. D. Vo, M. C. Mendel, S. G. Greenberg, J. Wang, D. F. Xia, J. C. Miller, F. D. Urnov, P. D. Gregory, and M. C. Holmes. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods*, 8(1):74–79, Jan. 2011.

S. J. Du, Z. Y. Gong, G. L. Fletcher, M. A. Shears, M. J. King, D. R. Idler, and C. L. Hew. Growth enhancement in transgenic Atlantic salmon by the use of an "all fish" chimeric growth hormone gene construct. *Biotechnology (N.Y.)*, 10(2):176–181, Feb. 1992.

Y. Du, J. Bi, J. Liu, X. Liu, X. Wu, P. Jiang, D. Yoo, Y. Zhang, J. Wu, R. Wan, X. Zhao, L. Guo, W. Sun, X. Cong, L. Chen, and J. Wang. 3cpro of foot-and-mouth disease virus antagonizes the interferon signaling pathway by blocking STAT1/STAT2 nuclear translocation. *J. Virol.*, 88(9):4908–4920, May 2014.

A. J. Dupuy, K. Clark, C. M. Carlson, S. Fritz, A. E. Davidson, K. M. Markley, K. Finley, C. F. Fletcher, S. C. Ekker, P. B. Hackett, S. Horn, and D. A. Largaespada. Mammalian germ-line transgenesis by transposition. *Proc. Natl. Acad. Sci. U.S.A.*, 99(7):4495–4499, Apr. 2002.

V. R. Edgerton, J. L. Smith, and D. R. Simpson. Muscle fibre type populations of human leg muscles. *Histochem. J.*, 7(3):259–266, May 1975.

H. M. Eilken, S.-I. Nishikawa, and T. Schroeder. Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature*, 457(7231):896–900, Feb. 2009.

B. Ekser, M. Ezzelarab, H. Hara, D. J. van der Windt, M. Wijkstrom, R. Bottino, M. Trucco, and D. K. C. Cooper. Clinical xenotransplantation: the next medical revolution? *Lancet*, 379(9816):672–683, Feb. 2012.

J.-C. Epinat, S. Arnould, P. Chames, P. Rochaix, D. Desfontaines, C. Puzin, A. Patin, A. Zanghellini, F. Paques, and E. Lacroix. A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, 31(11):2952–2962, June 2003.

P. Erickson, J. Gao, K. S. Chang, T. Look, E. Whisenant, S. Raimondi, R. Lasher, J. Trujillo, J. Rowley, and H. Drabkin. Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood*, 80(7):1825–1831, Oct. 1992.

F. Esashi, N. Christ, J. Gannon, Y. Liu, T. Hunt, M. Jasin, and S. C. West. CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature*, 434(7033):598–604, Mar. 2005.

M. J. Evans and M. H. Kaufman. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292(5819):154–156, July 1981.

M. Eyrich, S. C. Schreiber, G. Wollny, H. Ziegler, R. Schlenker, K. Koch-Buettner, M. Woelfl, P. G. Schlegel, and K. Schilbach. Pre-differentiated human committed T-lymphoid progenitors promote peripheral T-cell re-constitution after stem cell transplantation in immunodeficient mice. *Eur. J. Immunol.*, 41(12):3596–3603, Dec. 2011.

M. M. Falk, P. R. Grigera, I. E. Bergmann, A. Zibert, G. Multhaup, and E. Beck. Foot-and-mouth disease virus protease 3c induces specific proteolytic cleavage of host cell histone H3. *J. Virol.*, 64(2):748–756, Feb. 1990.

C. B. Fehilly, S. M. Willadsen, and E. M. Tucker. Interspecific chimaerism between sheep and goat. *Nature*, 307(5952):634–636, Feb. 1984.

S. Fekairi, S. Scaglione, C. Chahwan, E. R. Taylor, A. Tissier, S. Coulon, M.-Q. Dong, C. Ruse, J. R. Yates, P. Russell, R. P. Fuchs, C. H. McGowan, and P.-H. L. Gaillard. Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell*, 138(1):78–89, July 2009.

Q. Feng, M. A. Langereis, and F. J. M. van Kuppeveld. Induction and suppression of innate antiviral responses by picornaviruses. *Cytokine Growth Factor Rev.*, 25(5):577–585, Oct. 2014.

G. Ferrari, G. Cusella-De Angelis, M. Coletta, E. Paolucci, A. Stornaiuolo, G. Cossu, and

F. Mavilio. Muscle regeneration by bone marrow-derived myogenic progenitors. *Science*, 279(5356):1528–1530, Mar. 1998.

B. N. Fields, D. M. Knipe, and P. M. Howley, editors. *Fields virology*. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, 5th ed edition, 2007. ISBN 9780781760607.

A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391(6669):806–811, Feb. 1998.

K. E. Flick, M. S. Jurica, R. J. Monnat, Jr, and B. L. Stoddard. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, 394(6688):96–101, July 1998.

N. Foeger, W. Glaser, and T. Skern. Recognition of eukaryotic initiation factor 4g isoforms by picornaviral proteinases. *J. Biol. Chem.*, 277(46):44300–44309, Nov. 2002.

N. Foeger, E. Kuehnel, R. Cencic, and T. Skern. The binding of foot-and-mouth disease virus leader proteinase to eIF4gi involves conserved ionic interactions. *FEBS J.*, 272(10):2602–2611, May 2005.

G. Fox, N. R. Parry, P. V. Barnett, B. McGinn, D. J. Rowlands, and F. Brown. The cell attachment site on foot-and-mouth disease virus includes the amino acid sequence RGD (arginine-glycine-aspartic acid). *J. Gen. Virol.*, 70 (Pt 3):625–637, Mar. 1989.

D. Fraidenraich, E. Stillwell, E. Romero, D. Wilkes, K. Manova, C. T. Basson, and R. Benezra. Rescue of cardiac defects in id knockout embryos by injection of embryonic stem cells. *Science*, 306(5694):247–252, Oct. 2004.

E. E. Fry, S. M. Lea, T. Jackson, J. W. Newman, F. M. Ellard, W. E. Blakemore, R. Abu-Ghazaleh, A. Samuel, A. M. King, and D. I. Stuart. The structure and function of a foot-and-mouth disease virus-oligosaccharide receptor complex. *EMBO J.*, 18(3):543–554, Feb. 1999.

Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, and J. D. Sander. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, 31(9):822–826, Sept. 2013.

Y. Fu, J. D. Sander, D. Reyon, V. M. Cascio, and J. K. Joung. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature Biotechnology*, 32(3):279–284, Jan. 2014.

S. D. Fugmann, A. I. Lee, P. E. Shockett, I. J. Villey, and D. G. Schatz. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu. Rev. Immunol.*, 18: 495–527, 2000a.

S. D. Fugmann, I. J. Villey, L. M. Ptaszek, and D. G. Schatz. Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. *Mol. Cell*, 5(1):97–107, Jan. 2000b.

W. Fujii, K. Kawasaki, K. Sugiura, and K. Naito. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Res.*, 41(20):e187, Nov. 2013.

T. Furuta, H. Takemura, Z.-Y. Liao, G. J. Aune, C. Redon, O. A. Sedelnikova, D. R. Pilch, E. P. Rogakou, A. Celeste, H. T. Chen, A. Nussenzweig, M. I. Aladjem, W. M. Bonner, and Y. Pommier. Phosphorylation of histone H2ax and activation of Mre11, Rad50, and Nbs1 in response to replication-dependent DNA double-strand breaks induced by mammalian DNA topoisomerase I cleavage complexes. *J. Biol. Chem.*, 278(22):20303–20312, May 2003.

O. Gafni, L. Weinberger, A. A. Mansour, Y. S. Manor, E. Chomsky, D. Ben-Yosef, Y. Kalma, S. Viukov, I. Maza, A. Zviran, Y. Rais, Z. Shipony, Z. Mukamel, V. Krupalnik, M. Zerbib, S. Geula, I. Caspi, D. Schneir, T. Shwartz, S. Gilad, D. Amann-Zalcenstein, S. Benjamin, I. Amit, A. Tanay, R. Massarwa, N. Novershtern, and J. H. Hanna. Derivation of novel human ground state naive pluripotent stem cells. *Nature*, Oct. 2013.

T. Gaj, C. A. Gersbach, and C. F. Barbas, 3rd. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.*, 31(7):397–405, July 2013.

V. M. Gantz and E. Bier. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*, 348(6233):442–444, Apr. 2015.

R. L. Gardner. Mouse chimeras obtained by the injection of cells into the blastocyst. *Nature*, 220(5167):596–597, Nov. 1968.

R. L. Gardner and M. H. Johnson. Investigation of early mammalian development using interspecific chimaeras between rat and mouse. *Nature New Biol.*, 246(151):86–89, Nov. 1973.

R. L. Gardner and A. J. Munro. Successful construction of chimaeric rabbit. *Nature*, 250(462): 146–147, July 1974.

J. E. Garneau, M.-v. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux,

P. Horvath, A. H. MagadÃ¡n, and S. Moineau. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320):67–71, Nov. 2010.

P. J. Gerber, H. Steinfeld, B. Henderson, A. Mottet, C. Opio, J. Dijkman, A. Falcucci, and G. Tempio. *Tackling climate change through livestock: a global assessment of emissions and mitigation opportunities*. FAO, Rome, 2013. ISBN 9789251079201.

M. B. Gerstein, J. Rozowsky, K.-K. Yan, D. Wang, C. Cheng, J. B. Brown, C. A. Davis, L. Hillier, C. Sisu, J. J. Li, B. Pei, A. O. Harmanci, M. O. Duff, S. Djebali, R. P. Alexander, B. H. Alver, R. Auerbach, K. Bell, P. J. Bickel, M. E. Boeck, N. P. Boley, B. W. Booth, L. Cherbas, P. Cherbas, C. Di, A. Dobin, J. Drenkow, B. Ewing, G. Fang, M. Fastuca, E. A. Feingold, A. Frankish, G. Gao, P. J. Good, R. Guigo, A. Hammonds, J. Harrow, R. A. Hoskins, C. Howald, L. Hu, H. Huang, T. J. P. Hubbard, C. Huynh, S. Jha, D. Kasper, M. Kato, T. C. Kaufman, R. R. Kitchen, E. Ladewig, J. Lagarde, E. Lai, J. Leng, Z. Lu, M. MacCoss, G. May, R. McWhirter, G. Merrihew, D. M. Miller, A. Mortazavi, R. Murad, B. Oliver, S. Olson, P. J. Park, M. J. Pazin, N. Perrimon, D. Pervouchine, V. Reinke, A. Reymond, G. Robinson, A. Samsonova, G. I. Saunders, F. Schlesinger, A. Sethi, F. J. Slack, W. C. Spencer, M. H. Stoiber, P. Strasbourger, A. Tanzer, O. A. Thompson, K. H. Wan, G. Wang, H. Wang, K. L. Watkins, J. Wen, K. Wen, C. Xue, L. Yang, K. Yip, C. Zaleski, Y. Zhang, H. Zheng, S. E. Brenner, B. R. Graveley, S. E. Celniker, T. R. Gingeras, and R. Waterston. Comparative analysis of the transcriptome across distant species. *Nature*, 512 (7515):445–448, Aug. 2014.

A. M. Geurts, G. J. Cost, Y. Freyvert, B. Zeitler, J. C. Miller, V. M. Choi, S. S. Jenkins, A. Wood, X. Cui, X. Meng, A. Vincent, S. Lam, M. Michalkiewicz, R. Schilling, J. Foeckler, S. Kalloway, H. Weiler, S. MÃ©noret, I. Anegon, G. D. Davis, L. Zhang, E. J. Rebar, P. D. Gregory, F. D. Urnov, H. J. Jacob, and R. Buelow. Knockout rats via embryo microinjection of zinc-finger nucleases. *Science*, 325(5939):433, July 2009.

M. C. Ghozi, Y. Bernstein, V. Negreanu, D. Levanon, and Y. Groner. Expression of the human acute myeloid leukemia gene AML1 is regulated by two promoter regions. *Proc. Natl. Acad. Sci. U.S.A.*, 93(5):1935–1940, Mar. 1996.

B. Giannesini, C. Vilmen, H. Amthor, M. Bernard, and D. Bendahan. Lack of myostatin impairs mechanical performance and ATP cost of contraction in exercising mouse gastrocnemius muscle in vivo. *Am. J. Physiol. Endocrinol. Metab.*, 305(1):E33–40, July 2013.

L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, and J. S. Weissman. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*, Oct. 2014.

S. F. Gilbert. *Developmental biology*. Sinauer Associates, Sunderland, Mass, 9th ed edition, 2010.

S. Girgenrath, K. Song, and L.-A. Whittemore. Loss of myostatin expression alters fiber-type distribution and expression of myosin heavy chain isoforms in slow- and fast-type skeletal muscle. *Muscle Nerve*, 31(1):34–40, Jan. 2005.

W. Glaser and T. Skern. Extremely efficient cleavage of eIF4g by picornaviral proteinases L and 2a in vitro. *FEBS Lett.*, 480(2-3):151–155, Sept. 2000.

W. Glaser, R. Cencic, and T. Skern. Foot-and-mouth disease virus leader proteinase: involvement of C-terminal residues in self-processing and cleavage of eIF4gi. *J. Biol. Chem.*, 276 (38):35473–35481, Sept. 2001.

R. S. Goldstein, M. Drukker, B. E. Reubinoff, and N. Benvenisty. Integration and differentiation of human embryonic stem cells transplanted to the chick embryo. *Dev. Dyn.*, 225(1): 80–86, Sept. 2002.

S. P. Golovan, R. G. Meidinger, A. Ajakaiye, M. Cottrill, M. Z. Wiederkehr, D. J. Barney, C. Plante, J. W. Pollard, M. Z. Fan, M. A. Hayes, J. Laursen, J. P. Hjorth, R. R. Hacker, J. P. Phillips, and C. W. Forsberg. Pigs expressing salivary phytase produce low-phosphorus manure. *Nat. Biotechnol.*, 19(8):741–745, Aug. 2001.

E. I. Golub, R. C. Gupta, T. Haaf, M. S. Wold, and C. M. Radding. Interaction of human rad51 recombination protein with single-stranded DNA binding protein, RPA. *Nucleic Acids Res.*, 26(23):5388–5393, Dec. 1998.

J. W. Gordon and F. H. Ruddle. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science*, 214(4526):1244–1246, Dec. 1981.

A. Gradi, H. Imataka, Y. V. Svitkin, E. Rom, B. Raught, S. Morino, and N. Sonenberg. A novel functional human eukaryotic translation initiation factor 4g. *Mol. Cell. Biol.*, 18(1):334–342, Jan. 1998.

A. Gradi, N. Foeger, R. Strong, Y. V. Svitkin, N. Sonenberg, T. Skern, and G. J. Belsham. Cleavage of eukaryotic translation initiation factor 4gii within foot-and-mouth disease virus-

infected cells: identification of the L-protease cleavage site in vitro. *J. Virol.*, 78(7):3271–3278, Apr. 2004.

F. L. Graham and A. J. van der Eb. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*, 52(2):456–467, Apr. 1973.

U. Grawunder, M. Wilm, X. Wu, P. Kulesza, T. E. Wilson, M. Mann, and M. R. Lieber. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature*, 388(6641):492–495, July 1997.

S. Grizot, J. Smith, F. Daboussi, J. Prieto, P. Redondo, N. Merino, M. Villate, S. Thomas, L. Lemaire, G. Montoya, F. J. Blanco, F. Paques, and P. Duchateau. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.*, 37(16): 5405–5419, Sept. 2009.

L. Grobet, L. J. Martin, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, A. Schoeberlein, S. Dunner, F. MÃ©nissier, J. Massabanda, R. Fries, R. Hanset, and M. Georges. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat. Genet.*, 17 (1):71–74, Sept. 1997.

R. W. J. Groen, W. A. Noort, R. A. Raymakers, H.-J. Prins, L. Aalders, F. M. Hofhuis, P. Moerer, J. F. van Velzen, A. C. Bloem, B. van Kessel, H. Rozemuller, E. van Binsbergen, A. Buijs, H. Yuan, J. D. de Bruijn, M. de Weers, P. W. H. I. Parren, J. J. Schuringa, H. M. Lokhorst, T. Mutis, and A. C. M. Martens. Reconstructing the human hematopoietic niche in immune deficient mice, opportunities for studying primary multiple myeloma. *Blood*, May 2012. ISSN 1528-0020.

J. D. Growney, H. Shigematsu, Z. Li, B. H. Lee, J. Adelsperger, R. Rowan, D. P. Curley, J. L. Kutok, K. Akashi, I. R. Williams, N. A. Speck, and D. G. Gilliland. Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*, 106(2): 494–504, July 2005.

M. J. Grubman, M. Zellner, G. Bablanian, P. W. Mason, and M. E. Piccone. Identification of the active-site residues of the 3c proteinase of foot-and-mouth disease virus. *Virology*, 213 (2):581–589, Nov. 1995.

G. J. Grundy, W. Yang, and M. Gellert. Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. *Proc. Natl. Acad. Sci. U.S.A.*, 107(52):22487–22492, Dec. 2010.

J. Gu, H. Lu, B. Tippin, N. Shimazaki, M. F. Goodman, and M. R. Lieber. XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J.*, 26(4): 1010–1023, Feb. 2007.

A. Guarne, J. Tormo, R. Kirchweger, D. Pfistermueller, I. Fita, and T. Skern. Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4g recognition. *EMBO J.*, 17(24):7469–7479, Dec. 1998.

A. Guarne, B. Hampoelz, W. Glaser, X. Carpena, J. Tormo, I. Fita, and T. Skern. Structural and biochemical features distinguish the foot-and-mouth disease virus leader proteinase from other papain-like enzymes. *J. Mol. Biol.*, 302(5):1227–1240, Oct. 2000.

J. P. Guilinger, D. B. Thompson, and D. R. Liu. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.*, 32(6):577–582, June 2014.

J. Guirouilh-Barbat, E. Rass, I. Plo, P. Bertrand, and B. S. Lopez. Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. *Proc. Natl. Acad. Sci. U.S.A.*, 104(52):20902–20907, Dec. 2007.

J. Guiu, R. Shimizu, T. D'Altri, S. T. Fraser, J. Hatakeyama, E. H. Bresnick, R. Kageyama, E. Dzierzak, M. Yamamoto, L. Espinosa, and A. Bigas. Hes repressors are essential regulators of hematopoietic stem cell development downstream of Notch signaling. *J. Exp. Med.*, 210(1):71–84, Jan. 2013.

H. Guo, S. Zimmerly, P. S. Perlman, and A. M. Lambowitz. Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J.*, 16(22):6835–6848, Nov. 1997.

H. Guo, M. Karberg, M. Long, J. P. Jones, 3rd, B. Sullenger, and A. M. Lambowitz. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science*, 289(5478):452–457, July 2000.

H. Guo, O. Ma, N. A. Speck, and A. D. Friedman. Runx1 deletion or dominant inhibition reduces Cebpa transcription via conserved promoter and distal enhancer sites to favor monopoiesis over granulopoiesis. *Blood*, 119(19):4408–4418, May 2012.

J. Guo, T. Gaj, and C. F. Barbas. Directed evolution of an enhanced and highly efficient FokI cleavage domain for zinc finger nucleases. *J. Mol. Biol.*, 400(1):96–107, July 2010.

E. Gussoni, Y. Soneoka, C. D. Strickland, E. A. Buzney, M. K. Khan, A. F. Flint, L. M. Kunkel,

and R. C. Mulligan. Dystrophin expression in the mdx mouse restored by stem cell transplantation. *Nature*, 401(6751):390–394, Sept. 1999.

A. Haghighat, Y. Svitkin, I. Novoa, E. Kuechler, T. Skern, and N. Sonenberg. The eIF4g-eIF4e complex is the target for direct cleavage by the rhinovirus 2a proteinase. *J. Virol.*, 70(12): 8444–8450, Dec. 1996.

J. E. Hall and A. C. Guyton. *Guyton and Hall textbook of medical physiology*. Saunders/Elsevier, Philadelphia, Pa., 2011.

D. L. Hamilton and K. Abremski. Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites. *J. Mol. Biol.*, 178(2):481–486, Sept. 1984.

R. E. Hammer, V. G. Pursel, C. E. Rexroad, Jr, R. J. Wall, D. J. Bolt, K. M. Ebert, R. D. Palmiter, and R. L. Brinster. Production of transgenic rabbits, sheep and pigs by microinjection. *Nature*, 315(6021):680–683, June 1985.

J. Han, J. Zhang, L. Chen, B. Shen, J. Zhou, B. Hu, Y. Du, P. H. Tate, X. Huang, and W. Zhang. Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biol*, 11(7): 829–835, July 2014.

X. Han, M. Chen, F. Wang, M. Windrem, S. Wang, S. Shanz, Q. Xu, N. A. Oberheim, L. Bekar, S. Betstadt, A. J. Silva, T. Takano, S. A. Goldman, and M. Nedergaard. Forebrain engraftment by human glial progenitor cells enhances synaptic plasticity and learning in adult mice. *Cell Stem Cell*, 12(3):342–353, Mar. 2013.

M. A. Hatlen, L. Wang, and S. D. Nimer. AML1-ETO driven acute leukemia: insights into pathogenesis and potential therapeutic approaches. *Front Med*, 6(3):248–262, Sept. 2012.

J. Hauschild, B. Petersen, Y. Santiago, A.-L. Queisser, J. W. Carnwath, A. Lucas-Hahn, L. Zhang, X. Meng, P. D. Gregory, R. Schwinzer, G. J. Cost, and H. Niemann. Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.*, 108(29):12013–12017, July 2011.

F. E. M. Haynes, P. L. Greenwood, M. B. McDonagh, and V. H. Oddy. Myostatin allelic status interacts with level of nutrition to affect growth, composition, and myofiber characteristics of lambs. *J. Anim. Sci.*, 90(2):456–465, Feb. 2012.

A. Hendel, E. J. Fine, G. Bao, and M. H. Porteus. Quantifying on- and off-target genome editing. *Trends Biotechnol.*, 33(2):132–140, Feb. 2015.

A. Hennebry, C. Berry, V. Siriett, P. O'Callaghan, L. Chau, T. Watson, M. Sharma, and R. Kam-

badur. Myostatin regulates fiber-type composition of skeletal muscle by regulating MEF2 and MyoD gene expression. *Am. J. Physiol., Cell Physiol.*, 296(3):C525–534, Mar. 2009.

M. W. Hentze. eIF4g–A Multipurpose Ribosome Adapter? *Science*, 275(5299):500–501, Jan. 1997.

K. Herbers, J. Conrads-Strauch, and U. Bonas. Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature*, 356 (6365):172–174, Mar. 1992.

J. E. Hesse, M. R. Lieber, K. Mizuuchi, and M. Gellert. V(D)J recombination: a functional definition of the joining signals. *Genes Dev.*, 3(7):1053–1061, July 1989.

J. J. Hill, M. V. Davies, A. A. Pearson, J. H. Wang, R. M. Hewick, N. M. Wolfman, and Y. Qiu. The myostatin propeptide and the follistatin-related gene are inhibitory binding proteins of myostatin in normal serum. *J. Biol. Chem.*, 277(43):40735–40741, Oct. 2002.

A. Hinnen, J. B. Hicks, and G. R. Fink. Transformation of yeast. *Proc. Natl. Acad. Sci. U.S.A.*, 75(4):1929–1933, Apr. 1978.

S. Hiroyuki and K. Susumu. New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). *Gene*, 16(1-3):73–78, Dec. 1981.

T. Horii, Y. Arai, M. Yamazaki, S. Morita, M. Kimura, M. Itoh, Y. Abe, and I. Hatada. Validation of microinjection methods for generating knockout mice by CRISPR/Cas-mediated genome engineering. *Scientific Reports*, 4, Mar. 2014.

Z. Hou, Y. Zhang, N. E. Propson, S. E. Howden, L.-F. Chu, E. J. Sontheimer, and J. A. Thomson. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc. Natl. Acad. Sci. U.S.A.*, 110(39):15644–15649, Sept. 2013.

P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31(9):827–832, July 2013.

B. Huang, C. J. Schaeffer, Q. Li, and M. D. Tsai. Splase: a new class IIS zinc-finger restriction endonuclease with specificity for Sp1 binding sites. *J. Protein Chem.*, 15(5):481–489, July 1996.

G. Huang, K. Shigesada, K. Ito, H. J. Wee, T. Yokomizo, and Y. Ito. Dimerization

with PEBP2beta protects RUNX1/AML1 from ubiquitin-proteasome-mediated degradation. *EMBO J.*, 20(4):723–733, Feb. 2001.

G. Huang, P. Zhang, H. Hirai, S. Elf, X. Yan, Z. Chen, S. Koschmieder, Y. Okuno, T. Dayaram, J. D. Growney, R. A. Shivdasani, D. G. Gilliland, N. A. Speck, S. D. Nimer, and D. G. Tenen. PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis. *Nat. Genet.*, 40(1):51–60, Jan. 2008.

H. Huang, A. J. Woo, Z. Waldon, Y. Schindler, T. B. Moran, H. H. Zhu, G.-S. Feng, H. Steen, and A. B. Cantor. A Src family kinase-Shp2 axis controls RUNX1 activity in megakaryocyte and T-lymphocyte differentiation. *Genes Dev.*, 26(14):1587–1601, July 2012.

L. Huesser, M. P. Alves, N. Ruggli, and A. Summerfield. Identification of the role of RIG-I, MDA-5 and TLR3 in sensing RNA viruses in porcine epithelial cells using lentivirus-driven RNA interference. *Virus Res.*, 159(1):9–16, July 2011.

Z. Ibrahim, J. Busch, M. Awwad, R. Wagner, K. Wells, and D. K. C. Cooper. Selected physiologic compatibilities and incompatibilities between human and porcine organ systems. *Xenotransplantation*, 13(6):488–499, Nov. 2006.

M. Ichikawa, S. Goyama, T. Asai, M. Kawazu, M. Nakagawa, M. Takeshita, S. Chiba, S. Ogawa, and M. Kurokawa. AML1/Runx1 negatively regulates quiescent hematopoietic stem cells in adult hematopoiesis. *J. Immunol.*, 180(7):4402–4408, Apr. 2008.

S. C. Y. Ip, U. Rass, M. G. Blanco, H. R. Flynn, J. M. Skehel, and S. C. West. Identification of Holliday junction resolvases from humans and yeast. *Nature*, 456(7220):357–361, Nov. 2008.

G. Ira, A. Malkova, G. Liberi, M. Foiani, and J. E. Haber. Srs2 and Sgs1-Top3 suppress crossovers during double-strand break repair in yeast. *Cell*, 115(4):401–411, Nov. 2003.

Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, and A. Nakata. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.*, 169(12):5429–5433, Dec. 1987.

A. Isotani, H. Hatayama, K. Kaseda, M. Ikawa, and M. Okabe. Formation of a thymus from rat ES cells in xenogeneic nude mouse-rat ES chimeras. *Genes Cells*, 16(4):397–405, Apr. 2011.

A. Israel. The IKK complex, a central regulator of NF-kappaB activation. *Cold Spring Harb Perspect Biol*, 2(3):a000158, Mar. 2010.

K. Itakura, T. Hirose, R. Crea, A. D. Riggs, H. L. Heyneker, F. Bolivar, and H. W. Boyer. Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science*, 198(4321):1056–1063, Dec. 1977.

D. A. Jackson, R. H. Symons, and P. Berg. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, 69 (10):2904–2909, Oct. 1972.

T. Jackson, A. Sharma, R. A. Ghazaleh, W. E. Blakemore, F. M. Ellard, D. L. Simmons, J. W. Newman, D. I. Stuart, and A. M. King. Arginine-glycine-aspartic acid-specific binding by foot-and-mouth disease viruses to the purified integrin alpha(v)beta3 in vitro. *J. Virol.*, 71 (11):8357–8361, Nov. 1997.

T. Jackson, W. Blakemore, J. W. Newman, N. J. Knowles, A. P. Mould, M. J. Humphries, and A. M. King. Foot-and-mouth disease virus is a ligand for the high-affinity binding conformation of integrin alpha5beta1: influence of the leucine residue within the RGDL motif on selectivity of integrin binding. *J. Gen. Virol.*, 81(Pt 5):1383–1391, May 2000a.

T. Jackson, D. Sheppard, M. Denyer, W. Blakemore, and A. M. King. The epithelial integrin alphavbeta6 is a receptor for foot-and-mouth disease virus. *J. Virol.*, 74(11):4949–4956, June 2000b.

T. Jackson, S. Clark, S. Berryman, A. Burman, S. Cambier, D. Mu, S. Nishimura, and A. M. Q. King. Integrin alphavbeta8 functions as a receptor for foot-and-mouth disease virus: role of the beta-chain cytodomain in integrin-mediated infection. *J. Virol.*, 78(9):4533–4540, May 2004.

P. T. Jacobs, L. Cao, J. B. Samon, C. A. Kane, E. E. Hedblom, A. Bowcock, and J. C. Telfer. Runx transcription factors repress human and murine c-Myc expression in a DNA-binding and C-terminally dependent manner. *PLoS ONE*, 8(7):e69083, 2013.

A. Jacquier and B. Dujon. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, 41(2):383–394, June 1985.

R. Jaenisch and B. Mintz. Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 71(4):1250–1254, Apr. 1974.

D. James, S. A. Noggle, T. Swigut, and A. H. Brivanlou. Contribution of human embryonic stem cells to mouse blastocysts. *Dev. Biol.*, 295(1):90–102, July 2006.

L. Jansson and J. Larsson. W41/W41 blastocyst complementation: a system for genetic modeling of hematopoiesis. *Blood*, 115(1):47–50, Jan. 2010.

Y. Ji, W. Resch, E. Corbett, A. Yamane, R. Casellas, and D. G. Schatz. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell*, 141(3):419–431, Apr. 2010.

W. Jiang, D. Bikard, D. Cox, F. Zhang, and L. A. Marraffini. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.*, Jan. 2013.

M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337 (6096):816–821, Aug. 2012.

P. L. Johnson, K. G. Dodds, W. E. Bain, G. J. Greer, N. J. McLean, R. J. McLaren, S. M. Galloway, T. C. van Stijn, and J. C. McEwan. Investigations into the GDF8 g+6723g-A polymorphism in New Zealand Texel sheep. *J. Anim. Sci.*, 87(6):1856–1864, June 2009.

H. D. Jones. Regulatory uncertainty over genome editing. *Nature Plants*, 1(1):14011, Jan. 2015.

D. Joulia, H. Bernardi, V. Garandel, F. Rabenoelina, B. Vernus, and G. Cabello. Mechanisms involved in the inhibition of myoblast proliferation and differentiation by myostatin. *Exp. Cell Res.*, 286(2):263–275, June 2003.

M. S. Jurica, R. J. Monnat, Jr, and B. L. Stoddard. DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol. Cell*, 2(4):469–476, Oct. 1998.

E. B. Kabotyanski, L. Gomelsky, J. O. Han, T. D. Stamato, and D. B. Roth. Double-strand break repair in Ku86- and XRCC4-deficient cells. *Nucleic Acids Res.*, 26(23):5333–5342, Dec. 1998.

T. Kanno, Y. Kanno, L. F. Chen, E. Ogawa, W. Y. Kim, and Y. Ito. Intrinsic transcriptional activation-inhibition domains of the polyomavirus enhancer binding protein 2/core binding factor alpha subunit revealed in the presence of the beta subunit. *Mol. Cell. Biol.*, 18(5): 2444–2454, May 1998.

V. V. Kapitonov and J. Jurka. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.*, 3(6):e181, June 2005.

J. M. Karo, D. G. Schatz, and J. C. Sun. The RAG recombinase dictates functional heterogeneity and cellular fitness in natural killer cells. *Cell*, 159(1):94–107, Sept. 2014.

M. B. Kastan, Q. Zhan, W. S. el Deiry, F. Carrier, T. Jacks, W. V. Walsh, B. S. Plunkett, B. Vogelstein, and A. J. Fornace. A mammalian cell cycle checkpoint pathway utilizing p53 and GADD45 is defective in ataxia-telangiectasia. *Cell*, 71(4):587–597, Nov. 1992.

H. Kato, O. Takeuchi, S. Sato, M. Yoneyama, M. Yamamoto, K. Matsui, S. Uematsu, A. Jung, T. Kawai, K. J. Ishii, O. Yamaguchi, K. Otsu, T. Tsujimura, C.-S. Koh, C. Reis e Sousa, Y. Matsuura, T. Fujita, and S. Akira. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*, 441(7089):101–105, May 2006.

T. Kawai and S. Akira. Regulation of innate immune signalling pathways by the tripartite motif (TRIM) family proteins: TRIM regulation of innate immunity. *EMBO Molecular Medicine*, 3(9):513–527, Sept. 2011.

S. Kay, S. Hahn, E. Marois, G. Hause, and U. Bonas. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science*, 318(5850):648–651, Oct. 2007.

W. T. Khaled and P. Liu. Cancer mouse models: past, present and future. *Semin. Cell Dev. Biol.*, 27:54–60, Mar. 2014.

Y. Kim, J. Kweon, A. Kim, J. K. Chon, J. Y. Yoo, H. J. Kim, S. Kim, C. Lee, E. Jeong, E. Chung, D. Kim, M. S. Lee, E. M. Go, H. J. Song, H. Kim, N. Cho, D. Bang, S. Kim, and J.-S. Kim. A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.*, Feb. 2013a.

Y. Kim, J. Kweon, and J.-S. Kim. TALENs and ZFNs are associated with different mutation signatures. *Nat. Methods*, 10(3):185, Mar. 2013b.

Y. G. Kim and S. Chandrasegaran. Chimeric restriction endonuclease. *Proc. Natl. Acad. Sci. U.S.A.*, 91(3):883–887, Feb. 1994.

Y. G. Kim, J. Cha, and S. Chandrasegaran. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.*, 93(3):1156–1160, Feb. 1996.

Y.-K. Kim, G. Wee, J. Park, J. Kim, D. Baek, J.-S. Kim, and V. N. Kim. TALEN-based knockout library for human microRNAs. *Nat. Struct. Mol. Biol.*, Nov. 2013c.

M. Kinutani and N. M. Le Douarin. Avian spinal cord chimeras. I. Hatching ability and

posthatching survival in homo- and heterospecific chimeras. *Dev. Biol.*, 111(1):243–255, Sept. 1985.

R. Kirchweger, E. Ziegler, B. J. Lamphear, D. Waters, H. D. Liebig, W. Sommergruber, F. Sobrino, C. Hohenadl, D. Blaas, and R. E. Rhoads. Foot-and-mouth disease virus leader proteinase: purification of the Lb form and determination of its cleavage site on eIF-4 gamma. *J. Virol.*, 68(9):5677–5684, Sept. 1994.

K. Kissa and P. Herbomel. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature*, 464(7285):112–115, Mar. 2010.

R. P. Kitching. Clinical variation in foot and mouth disease: cattle. *Rev. - Off. Int. Epizoot.*, 21 (3):499–504, Dec. 2002.

R. P. Kitching and S. Alexandersen. Clinical variation in foot and mouth disease: pigs. *Rev. - Off. Int. Epizoot.*, 21(3):513–518, Dec. 2002.

R. P. Kitching and G. J. Hughes. Clinical variation in foot and mouth disease: sheep and goats. *Rev. - Off. Int. Epizoot.*, 21(3):505–512, Dec. 2002.

K. A. Klein and W. T. Jackson. Picornavirus subversion of the autophagy pathway. *Viruses*, 3 (9):1549–1561, Sept. 2011.

B. P. Kleinstiver, L. Wang, J. A. Wolfs, T. Kolaczyk, B. McDowell, X. Wang, C. Schild-Poulter, A. J. Bogdanove, and D. R. Edgell. The I-TevI Nuclease and Linker Domains Contribute to the Specificity of Monomeric TALENs. *G3 (Bethesda)*, Apr. 2014.

N. Klymiuk, B. Aigner, G. Brem, and E. Wolf. Genetic modification of pigs as organ donors for xenotransplantation. *Mol. Reprod. Dev.*, 77(3):209–221, Mar. 2010.

K. Knezevic, T. Bee, N. K. Wilson, M. E. Janes, S. Kinston, S. Polderdijk, A. Kolb-Kokocinski, K. Ottersbach, N. Pencovich, Y. Groner, M. de Bruijn, B. Goettgens, and J. E. Pimanda. A Runx1-Smad6 rheostat controls Runx1 activity during embryonic hematopoiesis. *Mol. Cell. Biol.*, 31(14):2817–2826, July 2011.

T. J. D. Knight-Jones and J. Rushton. The economic impacts of foot and mouth disease - what are they, how big are they and where do they occur? *Prev. Vet. Med.*, 112(3-4):161–173, Nov. 2013.

T. Kobayashi, T. Yamaguchi, S. Hamanaka, M. Kato-Itoh, Y. Yamazaki, M. Ibata, H. Sato, Y.-S. Lee, J.-I. Usui, A. S. Knisely, M. Hirabayashi, and H. Nakauchi. Generation of rat

pancreas in mouse by interspecific blastocyst injection of pluripotent stem cells. *Cell*, 142 (5):787–799, Sept. 2010.

T. Kobayashi, M. Kato-Itoh, and H. Nakauchi. Targeted Organ Generation Using Mixl1-Inducible Mouse Pluripotent Stem Cells in Blastocyst Complementation. *Stem Cells Dev.*, Oct. 2014.

H. Koike-Yusa, Y. Li, E.-P. Tan, M. D. C. Velasco-Herrera, and K. Yusa. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.*, 32(3):267–273, Mar. 2014.

F. O. Kok, M. Shin, C.-W. Ni, A. Gupta, A. S. Grosse, A. van Impel, B. C. Kirchmaier, J. Peterson-Maduro, G. Kourkoulis, I. Male, D. F. DeSantis, S. Sheppard-Tindell, L. Ebarasi, C. Betsholtz, S. Schulte-Merker, S. A. Wolfe, and N. D. Lawson. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell*, 32(1):97–108, Jan. 2015.

V. G. Kolupaeva, C. U. Hellen, and I. N. Shatsky. Structural analysis of the interaction of the pyrimidine tract-binding protein with the internal ribosomal entry site of encephalomyocarditis virus and foot-and-mouth disease virus RNAs. *RNA*, 2(12):1199–1212, Dec. 1996.

V. G. Kolupaeva, I. B. Lomakin, T. V. Pestova, and C. U. T. Hellen. Eukaryotic initiation factors 4g and 4a mediate conformational changes downstream of the initiation codon of the encephalomyocarditis virus internal ribosomal entry site. *Mol. Cell. Biol.*, 23(2):687–698, Jan. 2003.

Y. Komeno, M. Yan, S. Matsuura, K. Lam, M.-C. Lo, Y.-J. Huang, D. G. Tenen, J. R. Downing, and D.-E. Zhang. Runx1 exon 6-related alternative splicing isoforms differentially regulate hematopoiesis in mice. *Blood*, 123(24):3760–3769, June 2014.

T. Komori, H. Yagi, S. Nomura, A. Yamaguchi, K. Sasaki, K. Deguchi, Y. Shimizu, R. T. Bronson, Y. H. Gao, M. Inada, M. Sato, R. Okamoto, Y. Kitamura, S. Yoshiki, and T. Kishimoto. Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell*, 89(5):755–764, May 1997.

S. Konermann, M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg, H. Nishimasu, O. Nureki, and F. Zhang. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, Dec. 2014.

S. T. Kosak, J. A. Skok, K. L. Medina, R. Riblet, M. M. Le Beau, A. G. Fisher, and H. Singh. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, 296(5565):158–162, Apr. 2002.

M. Kurokawa, T. Tanaka, K. Tanaka, N. Hirano, S. Ogawa, K. Mitani, Y. Yazaki, and H. Hirai. A conserved cysteine residue in the runt homology domain of AML1 is required for the DNA binding ability and the transforming activity on fibroblasts. *J. Biol. Chem.*, 271(28): 16870–16876, July 1996.

J. Kwon, K. B. Morshead, J. R. Guyon, R. E. Kingston, and M. A. Oettinger. Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA. *Mol. Cell*, 6(5):1037–1048, Nov. 2000.

E. Lagasse, H. Connors, M. Al-Dhalimy, M. Reitsma, M. Dohse, L. Osborne, X. Wang, M. Finegold, I. L. Weissman, and M. Grompe. Purified hematopoietic stem cells can differentiate into hepatocytes in vivo. *Nat. Med.*, 6(11):1229–1234, Nov. 2000.

K. Lam, A. Muselman, R. Du, Y. Harada, A. G. Scholl, M. Yan, S. Matsuura, S. Weng, H. Harada, and D.-E. Zhang. Hmga2 is a direct target gene of RUNX1 and regulates expansion of myeloid progenitors in mice. *Blood*, 124(14):2203–2212, Oct. 2014.

C. Lancrin, P. Sroczynska, C. Stephenson, T. Allen, V. Kouskoff, and G. Lacaud. The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature*, 457(7231):892–895, Feb. 2009.

C. Lancrin, M. Mazan, M. Stefanska, R. Patel, M. Lichtinger, G. Costa, O. Vargel, N. K. Wilson, T. Moroy, C. Bonifer, B. Goettgens, V. Kouskoff, and G. Lacaud. GFI1 and GFI1b control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment. *Blood*, June 2012.

M. Landthaler, B. W. Shen, B. L. Stoddard, and D. A. Shub. I-BasI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities. *J. Mol. Biol.*, 358(4):1137–1151, May 2006.

P. Lawrence, E. A. Schafer, and E. Rieder. The nuclear protein Sam68 is cleaved by the FMDV 3c protease redistributing Sam68 to the cytoplasm during FMDV infection of host cells. *Virology*, 425(1):40–52, Mar. 2012.

N. Le Douarin. A biological cell labeling technique and its use in expermental embryology. *Dev. Biol.*, 30(1):217–222, Jan. 1973.

N. M. Le Douarin, D. Renaud, M. A. Teillet, and G. H. Le Douarin. Cholinergic differentiation of presumptive adrenergic neuroblasts in interspecific chimeras after heterotopic transplantations. *Proc. Natl. Acad. Sci. U.S.A.*, 72(2):728–732, Feb. 1975.

H. J. Lee, E. Kim, and J.-S. Kim. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.*, 20(1):81–89, Jan. 2010.

J.-H. Lee and T. T. Paull. Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science*, 304(5667):93–96, Apr. 2004.

J.-H. Lee and T. T. Paull. ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science*, 308(5721):551–554, Apr. 2005.

S. J. Lee and A. C. McPherron. Regulation of myostatin activity and muscle growth. *Proc. Natl. Acad. Sci. U.S.A.*, 98(16):9306–9311, July 2001.

Y. Lee, J. E. Manegold, A. D. Kim, C. Pouget, D. L. Stachura, W. K. Clements, and D. Traver. FGF signalling specifies haematopoietic stem cells through its regulation of somitic Notch signalling. *Nat Commun*, 5:5583, 2014.

Y.-S. Lee and S.-J. Lee. Regulation of GDF-11 and myostatin activity by GASP-1 and GASP-2. *Proc. Natl. Acad. Sci. U.S.A.*, 110(39):E3713–3722, Sept. 2013.

M. Lee-Theilen, A. J. Matthews, D. Kelly, S. Zheng, and J. Chaudhuri. CtIP promotes microhomology-mediated alternative end joining during class-switch recombination. *Nat. Struct. Mol. Biol.*, 18(1):75–79, Jan. 2011.

W. Lei, Q. Liang, L. Jing, C. Wang, X. Wu, and H. He. BoLA-DRB3 gene polymorphism and FMD resistance or susceptibility in Wanbei cattle. *Molecular Biology Reports*, 39(9): 9203–9209, Sept. 2012.

M. Leippert, E. Beck, F. Weiland, and E. Pfaff. Point mutations within the betaG-betaH loop of foot-and-mouth disease virus O1k affect virus attachment to target cells. *J. Virol.*, 71(2): 1046–1051, Feb. 1997.

C. Lemaitre, A. Grabarz, K. Tsouroula, L. Andronov, A. Furst, T. Pankotai, V. Heyer, M. Rogier, K. M. Attwood, P. Kessler, G. Dellaire, B. Klaholz, B. Reina-San-Martin, and E. Soutoglou. Nuclear position dictates DNA repair pathway choice. *Genes Dev.*, 28(22):2450–2463, Nov. 2014.

C. Lesterlin, G. Ball, L. Schermelleh, and D. J. Sherratt. RecA bundles mediate homology

pairing between distant sisters during DNA break repair. *Nature*, 506(7487):249–253, Feb. 2014.

T. M. Leu and D. G. Schatz. rag-1 and rag-2 are components of a high-molecular-weight complex, and association of rag-2 with this complex is rag-1 dependent. *Mol. Cell. Biol.*, 15 (10):5657–5670, Oct. 1995.

D. Levanon, G. Glusman, T. Bangsow, E. Ben-Asher, D. A. Male, N. Avidan, C. Bangsow, M. Hattori, T. D. Taylor, S. Taudien, K. Blechschmidt, N. Shimizu, A. Rosenthal, Y. Sakaki, D. Lancet, and Y. Groner. Architecture and anatomy of the genomic locus encoding the human leukemia-associated transcription factor RUNX1/AML1. *Gene*, 262(1-2):23–33, Jan. 2001.

D. Levanon, D. Bettoun, C. Harris-Cerruti, E. Woolf, V. Negreanu, R. Eilam, Y. Bernstein, D. Goldenberg, C. Xiao, M. Fliegauf, E. Kremer, F. Otto, O. Brenner, A. Lev-Tov, and Y. Groner. The Runx3 transcription factor regulates development and survival of TrkC dorsal root ganglia neurons. *EMBO J.*, 21(13):3454–3463, July 2002. ISSN 0261-4189. doi: 10.1093/emboj/cdf370.

E. Levantini, S. Lee, H. S. Radomska, C. J. Hetherington, M. Alberich-Jorda, G. Amabile, P. Zhang, D. A. Gonzalez, J. Zhang, D. S. Basseres, N. K. Wilson, S. Koschmieder, G. Huang, D.-E. Zhang, A. K. Ebralidze, C. Bonifer, Y. Okuno, B. Gottgens, and D. G. Tenen. RUNX1 regulates the CD34 gene in haematopoietic stem cells by mediating interactions with a distal regulatory element. *EMBO J.*, 30(19):4059–4070, Oct. 2011.

M. N. Levy, R. M. Berne, B. M. Koeppen, and B. A. Stanton, editors. *Berne & Levy principles of physiology*. Elsevier Mosby, St. Louis, Mo, 4th ed edition, 2006.

D. Li, Z. Qiu, Y. Shao, Y. Chen, Y. Guan, M. Liu, Y. Li, N. Gao, L. Wang, X. Lu, Y. Zhao, and M. Liu. Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat. Biotechnol.*, 31(8):681–683, Aug. 2013.

L. Li, L. P. Wu, and S. Chandrasegaran. Functional domains in Fok I restriction endonuclease. *Proc. Natl. Acad. Sci. U.S.A.*, 89(10):4275–4279, May 1992.

Q. Li, H. Imataka, S. Morino, G. W. Rogers, N. J. Richter-Cook, W. C. Merrick, and N. Sonenberg. Eukaryotic translation initiation factor 4aiii (eIF4aiii) is functionally distinct from eIF4ai and eIF4aii. *Mol. Cell. Biol.*, 19(11):7336–7346, Nov. 1999.

T. Li, S. Huang, W. Z. Jiang, D. Wright, M. H. Spalding, D. P. Weeks, and B. Yang. TAL

nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, 39(1):359–372, Jan. 2011.

W. Li, N. Ross-Smith, C. G. Proud, and G. J. Belsham. Cleavage of translation initiation factor 4ai (eIF4ai) but not eIF4aii by foot-and-mouth disease virus 3c protease: identification of the eIF4ai cleavage site. *FEBS Lett.*, 507(1):1–5, Oct. 2001.

Z. Li, M. J. Chen, T. Stacy, and N. A. Speck. Runx1 function in hematopoiesis is required in cells that express Tek. *Blood*, 107(1):106–110, Jan. 2006.

A. Liakhovitskaia, R. Gribi, E. Stamateris, G. Villain, T. Jaffredo, R. Wilkie, D. Gilchrist, J. Yang, J. Ure, and A. Medvinsky. Restoration of Runx1 expression in the Tie2 cell compartment rescues definitive hematopoietic stem cells and extends life of Runx1 knockout animals until birth. *Stem Cells*, 27(7):1616–1624, July 2009.

A. Liakhovitskaia, E. Lana-Elola, E. Stamateris, D. P. Rice, R. J. van 't Hof, and A. Medvinsky. The essential requirement for Runx1 in the development of the sternum. *Dev. Biol.*, 340(2): 539–546, Apr. 2010.

A. Liakhovitskaia, S. Rybtsov, T. Smith, A. Batsivari, N. Rybtsova, C. Rode, M. de Bruijn, F. Buchholz, S. Gordon-Keylock, S. Zhao, and A. Medvinsky. Runx1 is required for progression of CD41+ embryonic precursors into HSCs but not prior to this. *Development*, 141 (17):3319–3323, Sept. 2014.

F. Liang and M. Jasin. Ku80-deficient cells exhibit excess degradation of extrachromosomal DNA. *J. Biol. Chem.*, 271(24):14405–14411, June 1996.

M. Lichtinger, R. Ingram, R. Hannah, D. Mueller, D. Clarke, S. A. Assi, M. Lie-A-Ling, L. Noailles, M. S. Vijayabaskar, M. Wu, D. G. Tenen, D. R. Westhead, V. Kouskoff, G. Lacaud, B. Goettgens, and C. Bonifer. RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J.*, 31(22):4318–4333, Nov. 2012.

M. Lie-A-Ling, E. Marinopoulou, Y. Li, R. Patel, M. Stefanska, C. Bonifer, C. Miller, V. Kouskoff, and G. Lacaud. RUNX1 positively regulates a cell adhesion and migration program in murine hemogenic endothelium prior to blood emergence. *Blood*, 124(11):e11–20, Sept. 2014.

S. G. Lillico, A. Sherman, M. J. McGrew, C. D. Robertson, J. Smith, C. Haslam, P. Barnard, P. A. Radcliffe, K. A. Mitrophanous, E. A. Elliot, and H. M. Sang. Oviduct-specific expres-

sion of two therapeutic proteins in transgenic hens. *Proc. Natl. Acad. Sci. U.S.A.*, 104(6): 1771–1776, Feb. 2007.

S. G. Lillico, C. Proudfoot, D. F. Carlson, D. Stverakova, C. Neil, C. Blain, T. J. King, W. A. Ritchie, W. Tan, A. J. Mileham, D. G. McLaren, S. C. Fahrenkrug, and C. B. A. Whitelaw. Live pigs produced from genome edited zygotes. *Sci Rep*, 3:2847, 2013.

W. C. Lin and S. Desiderio. Cell cycle regulation of V(D)J recombination-activating protein RAG-2. *Proc. Natl. Acad. Sci. U.S.A.*, 91(7):2733–2737, Mar. 1994.

Y. Lin, E. J. Fine, Z. Zheng, C. J. Antico, R. A. Voit, M. H. Porteus, T. J. Cradick, and G. Bao. SAPTA: a new design tool for improving TALE nuclease activity. *Nucleic Acids Res.*, 42(6): e47, Apr. 2014.

M. Lisby, J. H. Barlow, R. C. Burgess, and R. Rothstein. Choreography of the DNA Damage Response. *Cell*, 118(6):699–713, Sept. 2004.

A. J. Little, E. Corbett, F. Ortega, and D. G. Schatz. Cooperative recruitment of HMGB1 during V(D)J recombination through interactions with RAG1 and DNA. *Nucleic Acids Res.*, Jan. 2013.

Q. Liu, S. Guntuku, X. S. Cui, S. Matsuoka, D. Cortez, K. Tamai, G. Luo, S. Carattini-Rivera, F. DeMayo, A. Bradley, L. A. Donehower, and S. J. Elledge. Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev.*, 14 (12):1448–1459, June 2000.

Y. Liu, D. Luo, H. Zhao, Z. Zhu, W. Hu, and C. H. K. Cheng. Inheritable and precise large genomic deletions of non-coding RNA genes in zebrafish using TALENs. *PLoS ONE*, 8(10): e76387, 2013.

G. Lozano, N. Fernandez, and E. Martinez-Salas. Magnesium-dependent folding of a picornavirus IRES element modulates RNA conformation and eIF4g interaction. *FEBS J.*, 281 (16):3685–3700, Aug. 2014.

M. Lusser, C. Parisi, D. Plan, and E. RodrÃguez-Cerezo. Deployment of new biotechnologies in plant breeding. *Nat. Biotechnol.*, 30(3):231–239, Mar. 2012.

J. Lyall, R. M. Irvine, A. Sherman, T. J. McKinley, A. Nunez, A. Purdie, L. Outtrim, I. H. Brown, G. Rolleston-Smith, H. Sang, and L. Tiley. Suppression of avian influenza transmission in genetically modified chickens. *Science*, 331(6014):223–226, Jan. 2011.

Y. Ma, U. Pannicke, K. Schwarz, and M. R. Lieber. Hairpin opening and overhang processing

by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell*, 108(6):781–794, Mar. 2002.

Y. Ma, H. Lu, B. Tippin, M. F. Goodman, N. Shimazaki, O. Koiwai, C.-L. Hsieh, K. Schwarz, and M. R. Lieber. A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol. Cell*, 16(5):701–713, Dec. 2004.

N. Maclean and S. Talwar. Injection of cloned genes into rainbow trout eggs. *Journal of Embryology and Experimental Morphology*, 82:187, 1984.

D. Maddalo, E. Manchado, C. P. Concepcion, C. Bonetti, J. A. Vidigal, Y.-C. Han, P. Ogrodowski, A. Crippa, N. Rekhtman, E. de Stanchina, S. W. Lowe, and A. Ventura. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature*, Oct. 2014.

S. Mader, H. Lee, A. Pause, and N. Sonenberg. The translation initiation factor eIF-4e binds to a common motif shared by the translation factor eIF-4 gamma and the translational repressors 4e-binding proteins. *Mol. Cell. Biol.*, 15(9):4990–4997, Sept. 1995.

B. W. J. Mahy. *Foot-and-mouth disease virus with 16 figures, mostly in color*. Springer, Berlin; New York, 2005. ISBN 9783540271093 3540271090. URL `http://public.eblib.com/choice/publicfullrecord.aspx?p=304001`.

A. N.-S. Mak, P. Bradley, R. A. Cernadas, A. J. Bogdanove, and B. L. Stoddard. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science*, 335(6069):716–719, Feb. 2012.

K. S. Makarova, D. H. Haft, R. Barrangou, S. J. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. M. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost, and E. V. Koonin. Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.*, 9(6):467–477, June 2011.

V. O. Mallet, C. Mitchell, E. Mezey, M. Fabre, J.-E. Guidotti, L. Renia, L. Coulombel, A. Kahn, and H. Gilgenkrantz. Bone marrow transplantation in mice leads to a minor population of hepatocytes that can be selectively amplified in vivo. *Hepatology*, 35(4):799–804, Apr. 2002.

P.-O. Mari, B. I. Florea, S. P. Persengiev, N. S. Verkaik, H. T. BrÃŒggenwirth, M. Modesti, G. Giglia-Mari, K. Bezstarosti, J. A. A. Demmers, T. M. Luider, A. B. Houtsmuller, and D. C. van Gent. Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc. Natl. Acad. Sci. U.S.A.*, 103(49):18597–18602, Dec. 2006.

G. R. Martin. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, 78(12):7634–7638, Dec. 1981.

A. Y. Masri, N. R. Lambe, J. M. Macfarlane, S. Brotherstone, W. Haresign, and L. BÃŒnger. Evaluating the effects of a single copy of a mutation in the myostatin gene (c.*1232g>A) on carcass traits in crossbred lambs. *Meat Sci.*, 87(4):412–418, Apr. 2011.

A. Matsakas, E. Mouisel, H. Amthor, and K. Patel. Myostatin knockout mice increase oxidative muscle phenotype as an adaptive response to exercise. *J. Muscle Res. Cell. Motil.*, 31(2): 111–125, Aug. 2010.

H. Matsunari, H. Nagashima, M. Watanabe, K. Umeyama, K. Nakano, M. Nagaya, T. Kobayashi, T. Yamaguchi, R. Sumazaki, L. A. Herzenberg, and H. Nakauchi. Blastocyst complementation generates exogenic pancreas in vivo in apancreatic cloned pigs. *Proc. Natl. Acad. Sci. U.S.A.*, 110(12):4557–4562, Mar. 2013.

S. Matsuoka, G. Rotman, A. Ogawa, Y. Shiloh, K. Tamai, and S. J. Elledge. Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proc. Natl. Acad. Sci. U.S.A.*, 97(19):10389–10394, Sept. 2000.

A. G. W. Matthews, A. J. Kuo, S. Ramon-Maiques, S. Han, K. S. Champagne, D. Ivanov, M. Gallardo, D. Carney, P. Cheung, D. N. Ciccone, K. L. Walter, P. J. Utz, Y. Shi, T. G. Kutateladze, W. Yang, O. Gozani, and M. A. Oettinger. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature*, 450(7172):1106–1110, Dec. 2007.

A. Mauro. Satellite cell of skeletal muscle fibers. *J Biophys Biochem Cytol*, 9:493–495, Feb. 1961.

A. Maxmen. Transgenic fish swims up regulatory stream. *Nature*, Dec. 2012.

C. Mayer, D. Neubauer, A. T. Nchinda, R. Cencic, K. Trompf, and T. Skern. Residue L143 of the foot-and-mouth disease virus leader proteinase is a determinant of cleavage specificity. *J. Virol.*, 82(9):4656–4659, May 2008.

J. F. Mayer, Jr and H. I. Fritz. The culture of preimplantation rat embryos and the production of allophenic rats. *J. Reprod. Fertil.*, 39(1):1–9, July 1974.

J. F. McBlane, D. C. van Gent, D. A. Ramsden, C. Romeo, C. A. Cuomo, M. Gellert, and

M. A. Oettinger. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell*, 83(3):387–395, Nov. 1995.

A. McConnell Smith, R. Takeuchi, S. Pellenz, L. Davis, N. Maizels, R. J. Monnat, Jr, and B. L. Stoddard. Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. U.S.A.*, 106(13): 5099–5104, Mar. 2009.

S. McCroskery, M. Thomas, L. Maxwell, M. Sharma, and R. Kambadur. Myostatin negatively regulates satellite cell activation and self-renewal. *J. Cell Biol.*, 162(6):1135–1147, Sept. 2003.

J. M. McCune, R. Namikawa, H. Kaneshima, L. D. Shultz, M. Lieberman, and I. L. Weissman. The SCID-hu mouse: murine model for the analysis of human hematolymphoid differentiation and function. *Science*, 241(4873):1632–1639, Sept. 1988.

L. J. McGuffin, K. Bryson, and D. T. Jones. The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4):404–405, Apr. 2000.

A. C. McPherron and S. J. Lee. Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl. Acad. Sci. U.S.A.*, 94(23):12457–12461, Nov. 1997.

A. C. McPherron and S.-J. Lee. Suppression of body fat accumulation in myostatin-deficient mice. *J. Clin. Invest.*, 109(5):595–601, Mar. 2002.

A. C. McPherron, A. M. Lawler, and S. J. Lee. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature*, 387(6628):83–90, May 1997.

A. C. McPherron, A. M. Lawler, and S. J. Lee. Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. *Nat. Genet.*, 22(3):260–264, July 1999.

M. Medina, E. Domingo, J. K. Brangwyn, and G. J. Belsham. The two species of the foot-and-mouth disease virus leader protein, expressed individually, exhibit the same activities. *Virology*, 194(1):355–359, May 1993.

A. Medvinsky and E. Dzierzak. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, 86(6):897–906, Sept. 1996.

A. Mendelson and P. S. Frenette. Hematopoietic stem cell niche maintenance during homeostasis and regeneration. *Nat. Med.*, 20(8):833–846, Aug. 2014.

S. Menoret, S. Fontaniere, D. Jantz, L. Tesson, R. Thinard, S. Remy, C. Usal, L.-H. Ouisse,

A. Fraichard, and I. Anegon. Generation of Rag1-knockout immunodeficient rats and mice using engineered meganucleases. *FASEB J.*, 27(2):703–711, Feb. 2013.

D. F. Mercer, D. E. Schiller, J. F. Elliott, D. N. Douglas, C. Hao, A. Rinfret, W. R. Addison, K. P. Fischer, T. A. Churchill, J. R. Lakey, D. L. Tyrrell, and N. M. Kneteman. Hepatitis C virus replication in mice with chimeric human livers. *Nat. Med.*, 7(8):927–933, Aug. 2001.

S. Meyers, J. R. Downing, and S. W. Hiebert. Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: the runt homology domain is required for DNA binding and protein-protein interactions. *Mol. Cell. Biol.*, 13 (10):6336–6345, Oct. 1993.

H. Mi, A. Muruganujan, J. T. Casagrande, and P. D. Thomas. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, 8(8):1551–1566, July 2013.

J. Miller, A. D. McLachlan, and A. Klug. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.*, 4(6):1609–1614, June 1985.

J. C. Miller, M. C. Holmes, J. Wang, D. Y. Guschin, Y.-L. Lee, I. Rupniewski, C. M. Beausejour, A. J. Waite, N. S. Wang, K. A. Kim, P. D. Gregory, C. O. Pabo, and E. J. Rebar. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.*, 25(7):778–785, July 2007.

H. Miyoshi, K. Shimizu, T. Kozu, N. Maseki, Y. Kaneko, and M. Ohki. t(8;21) breakpoints on chromosome 21 in acute myeloid leukemia are clustered within a limited region of a single gene, AML1. *Proc. Natl. Acad. Sci. U.S.A.*, 88(23):10431–10434, Dec. 1991.

H. Miyoshi, M. Ohira, K. Shimizu, K. Mitani, H. Hirai, T. Imai, K. Yokoyama, E. Soeda, and M. Ohki. Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. *Nucleic Acids Res.*, 23(14):2762–2769, July 1995.

E. A. Moehle, E. A. Moehle, J. M. Rock, J. M. Rock, Y.-L. Lee, Y. L. Lee, Y. Jouvenot, Y. Jouvenot, R. C. DeKelver, R. C. Dekelver, P. D. Gregory, P. D. Gregory, F. D. Urnov, F. D. Urnov, M. C. Holmes, and M. C. Holmes. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.*, 104 (9):3055–3060, Feb. 2007.

P. Mombaerts, J. Iacomini, R. S. Johnson, K. Herrup, S. Tonegawa, and V. E. Papaioannou.

RAG-1-deficient mice have no mature B and T lymphocytes. *Cell*, 68(5):869–877, Mar. 1992.

P. Monaghan, S. Gold, J. Simpson, Z. Zhang, P. H. Weinreb, S. M. Violette, S. Alexandersen, and T. Jackson. The alpha(v)beta6 integrin receptor for Foot-and-mouth disease virus is expressed constitutively on the epithelial cells targeted in cattle. *J. Gen. Virol.*, 86(Pt 10): 2769–2780, Oct. 2005.

A. F. Moon, M. Garcia-Diaz, V. K. Batra, W. A. Beard, K. Bebenek, T. A. Kunkel, S. H. Wilson, and L. C. Pedersen. The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair (Amst.)*, 6(12):1709–1725, Dec. 2007.

P. Moral-Lopez, E. Alvarez, N. Redondo, T. Skern, and L. Carrasco. L protease from foot and mouth disease virus confers eIF2-independent translation for mRNAs bearing picornavirus IRES. *FEBS Lett.*, 588(21):4053–4059, Nov. 2014.

C. S. Morgen and T. I. A. Sorensen. Obesity: global trends in the prevalence of overweight and obesity. *Nat Rev Endocrinol*, 10(9):513–514, Sept. 2014.

S. Morino, H. Imataka, Y. V. Svitkin, T. V. Pestova, and N. Sonenberg. Eukaryotic translation initiation factor 4e (eIF4e) binding site and the middle one-third of eIF4gi constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Mol. Cell. Biol.*, 20(2):468–477, Jan. 2000.

C. A. Morris. A review of genetic resistance to disease in Bos taurus cattle. *Vet. J.*, 174(3): 481–491, Nov. 2007.

J. F. Morrow, S. N. Cohen, A. C. Chang, H. W. Boyer, H. M. Goodman, and R. B. Helling. Replication and transcription of eukaryotic DNA in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.*, 71(5):1743–1747, May 1974.

M. J. Moscou and A. J. Bogdanove. A simple cipher governs DNA recognition by TAL effectors. *Science*, 326(5959):1501, Dec. 2009.

D. S. Mosher, P. Quignon, C. D. Bustamante, N. B. Sutter, C. S. Mellersh, H. G. Parker, and E. A. Ostrander. A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.*, 3(5):e79, May 2007.

E. Mouisel, K. Relizani, L. Mille-Hamard, R. Denis, C. HourdÃ©, O. Agbulut, K. Patel, L. Arandel, S. Morales-Gonzalez, A. Vignaud, L. Garcia, A. Ferry, S. Luquet, V. Billat, R. Ventura-Clapier, M. Schuelke, and H. Amthor. Myostatin is a key mediator between en-

ergy metabolism and endurance capacity of skeletal muscle. *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, 307(4):R444–454, Aug. 2014.

C. M. Moure, F. S. Gimble, and F. A. Quiocho. The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol.*, 334 (4):685–695, Dec. 2003.

A. M. Mueller, A. Medvinsky, J. Strouboulis, F. Grosveld, and E. Dzierzak. Development of hematopoietic stem cell activity in the mouse embryo. *Immunity*, 1(4):291–301, July 1994.

S. M. Mueller, G. Terszowski, C. Blum, C. Haller, V. Anquez, S. Kuschert, P. Carmeliet, H. G. Augustin, and H.-R. Rodewald. Gene targeting of VEGF-A in thymus epithelium disrupts thymus blood vessel architecture. *Proc. Natl. Acad. Sci. U.S.A.*, 102(30):10587–10592, July 2005.

I. M. Munoz, K. Hain, A.-C. Declais, M. Gardiner, G. W. Toh, L. Sanchez-Pulido, J. M. Heuckmann, R. Toth, T. Macartney, B. Eppink, R. Kanaar, C. P. Ponting, D. M. Lilley, and J. Rouse. Coordination of Structure-Specific Nucleases by Human SLX4/BTBD12 Is Required for DNA Repair. *Molecular Cell*, 35(1):116–127, July 2009.

A. R. Muotri, K. Nakashima, N. Toni, V. M. Sandler, and F. H. Gage. Development of functional human embryonic stem cell-derived neurons in mouse brain. *Proc. Natl. Acad. Sci. U.S.A.*, 102(51):18644–18648, Dec. 2005.

A. Nagamachi, P. W. Htun, F. Ma, K. Miyazaki, N. Yamasaki, M. Kanno, T. Inaba, Z.-i. Honda, T. Okuda, H. Oda, K. Tsuji, and H. Honda. A 5' untranslated region containing the IRES element in the Runx1 gene is required for angiogenesis, hematopoiesis and leukemogenesis in a knock-in mouse model. *Dev. Biol.*, 345(2):226–236, Sept. 2010.

M. Nakagawa, M. Ichikawa, K. Kumano, S. Goyama, M. Kawazu, T. Asai, S. Ogawa, M. Kurokawa, and S. Chiba. AML1/Runx1 rescues Notch1-null mutation-induced deficiency of para-aortic splanchnopleural hematopoiesis. *Blood*, 108(10):3329–3334, Nov. 2006.

K. Nakajima and Y. Yaoita. Comparison of TALEN scaffolds in Xenopus tropicalis. *Biol Open*, 2(12):1364–1370, 2013.

A. Nakata, M. Amemura, and K. Makino. Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *J. Bacteriol.*, 171(6):3553–3556, June 1989.

N. Nassif, J. Penney, S. Pal, W. R. Engels, and G. B. Gloor. Efficient copying of nonhomolog-

ous sequences from ectopic sites via P-element-induced gap repair. *Mol. Cell. Biol.*, 14(3): 1613–1625, Mar. 1994.

S. Neff, D. SÃ¡-Carvalho, E. Rieder, P. W. Mason, S. D. Blystone, E. J. Brown, and B. Baxt. Foot-and-mouth disease virus virulent for cattle utilizes the integrin alpha(v)beta3 as its receptor. *J. Virol.*, 72(5):3587–3594, May 1998.

W. Ni, J. Qiao, S. Hu, X. Zhao, M. Regouski, M. Yang, I. A. Polejaeva, and C. Chen. Efficient Gene Knockout in Goats Using CRISPR/Cas9 System. *PLoS ONE*, 9(9):e106718, 2014.

J. Nichols and A. Smith. Naive and primed pluripotent states. *Cell Stem Cell*, 4(6):487–492, June 2009.

B. Niebuhr, N. Kriebitzsch, M. Fischer, K. Behrens, T. GÃŒnther, M. Alawi, U. Bergholz, U. MÃŒller, S. Roscher, M. Ziegler, F. Buchholz, A. Grundhoff, and C. Stocking. Runx1 is essential at two stages of early murine B-cell development. *Blood*, 122(3):413–423, July 2013.

P. J. Nielsen and H. Trachsel. The mouse protein synthesis initiation factor 4a gene family includes two related functional genes which are differentially expressed. *EMBO J.*, 7(7): 2097–2105, July 1988.

H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, and O. Nureki. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156(5):935–949, Feb. 2014.

Y. Niu, K. Tenney, H. Li, and F. S. Gimble. Engineering variants of the I-SceI homing endonuclease with strand-specific and site-specific DNA-nicking activity. *J. Mol. Biol.*, 382(1): 188–202, Sept. 2008.

Y. Niu, B. Shen, Y. Cui, Y. Chen, J. Wang, L. Wang, Y. Kang, X. Zhao, W. Si, W. Li, A. P. Xiang, J. Zhou, X. Guo, Y. Bi, C. Si, B. Hu, G. Dong, H. Wang, Z. Zhou, T. Li, T. Tan, X. Pu, F. Wang, S. Ji, Q. Zhou, X. Huang, W. Ji, and J. Sha. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*, 156 (4):836–843, Feb. 2014.

T. North, T. L. Gu, T. Stacy, Q. Wang, L. Howard, M. Binder, M. MarÃn-Padilla, and N. A. Speck. Cbfa2 is required for the formation of intra-aortic hematopoietic clusters. *Development*, 126(11):2563–2575, June 1999.

T. E. North, M. F. T. R. de Bruijn, T. Stacy, L. Talebian, E. Lind, C. Robin, M. Binder,

E. Dzierzak, and N. A. Speck. Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity*, 16(5):661–672, May 2002.

W. T. Nottingham, A. Jarratt, M. Burgess, C. L. Speck, J.-F. Cheng, S. Prabhakar, E. M. Rubin, P.-S. Li, J. Sloane-Stanley, J. Kong-A-San, and M. F. T. R. de Bruijn. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood*, 110(13):4188–4197, Dec. 2007.

T. Ochi, A. N. Blackford, J. Coates, S. Jhujh, S. Mehmood, N. Tamura, J. Travers, Q. Wu, V. M. Draviam, C. V. Robinson, T. L. Blundell, and S. P. Jackson. DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science*, 347(6218):185–188, Jan. 2015.

M. A. Oettinger, D. G. Schatz, C. Gorka, and D. Baltimore. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science*, 248(4962):1517–1523, June 1990.

E. Ogawa, M. Inuzuka, M. Maruyama, M. Satake, M. Naito-Fujimoto, Y. Ito, and K. Shigesada. Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2 alpha. *Virology*, 194(1):314–331, May 1993.

B. M. Ogle, B. E. Knudsen, R. Nishitai, K. Ogata, and J. L. Platt. Toward development and production of human T cells in swine for potential use in adoptive T cell immunotherapy. *Tissue Eng Part A*, 15(5):1031–1040, May 2009.

Y. Ohi, H. Qin, C. Hong, L. Blouin, J. M. Polo, T. Guo, Z. Qi, S. L. Downey, P. D. Manos, D. J. Rossi, J. Yu, M. Hebrok, K. Hochedlinger, J. F. Costello, J. S. Song, and M. Ramalho-Santos. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nat. Cell Biol.*, 13(5):541–549, May 2011.

T. Ohlmann, V. M. Pain, W. Wood, M. Rau, and S. J. Morley. The proteolytic cleavage of eukaryotic initiation factor (eIF) 4g is prevented by eIF4e binding protein (PHAS-I; 4e-BP1) in the reticulocyte lysate. *EMBO J.*, 16(4):844–855, Feb. 1997.

H. Okada, T. Watanabe, M. Niki, H. Takano, N. Chiba, N. Yanai, K. Tani, H. Hibino, S. Asano, M. L. Mucenski, Y. Ito, T. Noda, and M. Satake. AML1(-/-) embryos do not express certain hematopoiesis-related gene transcripts including those of the PU.1 gene. *Oncogene*, 17(18): 2287–2293, Nov. 1998.

K. Okita, T. Ichisaka, and S. Yamanaka. Generation of germline-competent induced pluripotent stem cells. *Nature*, 448(7151):313–317, July 2007.

T. Okuda, J. van Deursen, S. W. Hiebert, G. Grosveld, and J. R. Downing. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, 84(2):321–330, Jan. 1996.

A. L. Olsen, D. L. Stachura, and M. J. Weiss. Designer blood: creating hematopoietic lineages from embryonic stem cells. *Blood*, 107(4):1265–1275, Feb. 2006.

C.-C. Pai, R. S. Deegan, L. Subramanian, C. Gal, S. Sarkar, E. J. Blaikley, C. Walker, L. Hulme, E. Bernhard, S. Codlin, J. Baehler, R. Allshire, S. Whitehall, and T. C. Humphrey. A histone H3k36 chromatin switch coordinates DNA double-strand break repair pathway choice. *Nat Commun*, 5:4091, 2014.

S. Parida. Vaccination against foot-and-mouth disease virus: strategies and effectiveness. *Expert Rev Vaccines*, 8(3):347–365, Mar. 2009.

K.-E. Park and B. P. V. L. Telugu. Role of stem cells in large animal genetic engineering in the TALENs-CRISPR era. *Reprod. Fertil. Dev.*, 26(1):65–73, 2013.

T. S. Park, H. J. Lee, K. H. Kim, J.-S. Kim, and J. Y. Han. Targeted gene knockout in chickens mediated by TALENs. *Proc. Natl. Acad. Sci. U.S.A.*, 111(35):12716–12721, Sept. 2014.

V. Pattanayak, S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna, and D. R. Liu. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology*, 31(9):839–843, Aug. 2013.

A. Pause and N. Sonenberg. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4a. *EMBO J.*, 11(7):2643–2654, July 1992.

N. P. Pavletich and C. O. Pabo. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, 252(5007):809–817, May 1991.

T. Pearson, D. L. Greiner, and L. D. Shultz. Humanized SCID mouse models for biomedical research. *Curr. Top. Microbiol. Immunol.*, 324:25–51, 2008.

B. Pessac, M. A. Bara, D. Ford, G. K. Patibandla, and D. Trisler. Hematopoietic progenitors express embryonic stem cell and germ layer genes. *C. R. Biol.*, 334(4):300–306, Apr. 2011.

B. Pessac, V. K. Nimmagadda, T. Makar, P. S. Fishman, C. T. Bever, and D. Trisler. Adult hematopoietic progenitors are multipotent in chimeric mice. *C. R. Biol.*, 335(7):454–462, July 2012.

T. V. Pestova, C. U. Hellen, and I. N. Shatsky. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol. Cell. Biol.*, 16(12):6859–6869, Dec. 1996a.

T. V. Pestova, I. N. Shatsky, and C. U. Hellen. Functional dissection of eukaryotic initiation factor 4f: the 4a subunit and the central domain of the 4g subunit are sufficient to mediate internal entry of 43s preinitiation complexes. *Mol. Cell. Biol.*, 16(12):6870–6878, Dec. 1996b.

J. B. Peter, R. J. Barnard, V. R. Edgerton, C. A. Gillespie, and K. E. Stempel. Metabolic profiles of three fiber types of skeletal muscle in guinea pigs and rabbits. *Biochemistry*, 11 (14):2627–2633, July 1972.

L. F. Peterson, A. Boyapati, V. Ranganathan, A. Iwama, D. G. Tenen, S. Tsai, and D.-E. Zhang. The Hematopoietic Transcription Factor AML1 (RUNX1) Is Negatively Regulated by the Cell Cycle Protein Cyclin D3. *Molecular and Cellular Biology*, 25(23):10205–10219, Dec. 2005.

D. Pette and R. S. Staron. Myosin isoforms, muscle fiber types, and transitions. *Microsc. Res. Tech.*, 50(6):500–509, Sept. 2000.

B. Philip, Z. Lu, and Y. Gao. Regulation of GDF-8 signaling by the p38 MAPK. *Cell. Signal.*, 17(3):365–375, Mar. 2005.

E. V. Pilipenko, T. V. Pestova, V. G. Kolupaeva, E. V. Khitrina, A. N. Poperechnaya, V. I. Agol, and C. U. Hellen. A cell cycle-dependent protein serves as a template-specific translation initiation factor. *Genes Dev.*, 14(16):2028–2045, Aug. 2000.

D. Pineiro, J. Ramajo, S. S. Bradrick, and E. Martinez-Salas. Gemin5 proteolysis reveals a novel motif to identify L protease targets. *Nucleic Acids Res.*, 40(11):4942–4953, June 2012.

R. J. Platt, S. Chen, Y. Zhou, M. J. Yim, L. Swiech, H. R. Kempton, J. E. Dahlman, O. Parnas, T. M. Eisenhaure, M. Jovanovic, D. B. Graham, S. Jhunjhunwala, M. Heidenreich, R. J. Xavier, R. Langer, D. G. Anderson, N. Hacohen, A. Regev, G. Feng, P. A. Sharp, and F. Zhang. CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. *Cell*, 159(2):440–455, Oct. 2014.

B. Pohajdak, M. Mansour, O. Hrytsenko, J. M. Conlon, L. C. Dymond, and J. R. Wright,

Jr. Production of transgenic tilapia with Brockmann bodies secreting [desThrB30] human insulin. *Transgenic Res.*, 13(4):313–323, Aug. 2004.

S. E. Polo, A. Kaidi, L. Baskcomb, Y. Galanty, and S. P. Jackson. Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *EMBO J.*, 29(18):3130–3139, Sept. 2010.

C. Porta, X. Xu, S. Loureiro, S. Paramasivam, J. Ren, T. Al-Khalil, A. Burman, T. Jackson, G. J. Belsham, S. Curry, G. P. Lomonossoff, S. Parida, D. Paton, Y. Li, G. Wilsden, N. Ferris, R. Owens, A. Kotecha, E. Fry, D. I. Stuart, B. Charleston, and I. M. Jones. Efficient production of foot-and-mouth disease virus empty capsids in insect cells following down regulation of 3c protease activity. *J. Virol. Methods*, 187(2):406–412, Feb. 2013. ISSN 1879-0984.

M. H. Porteus and D. Baltimore. Chimeric nucleases stimulate gene targeting in human cells. *Science*, 300(5620):763, May 2003.

A. J. Potocnik, H. Kohler, and K. Eichmann. Hemato-lymphoid in vivo reconstitution potential of subpopulations derived from in vitro differentiated embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, 94(19):10295–10300, Sept. 1997.

W. T. Poueymirou, W. Auerbach, D. Frendewey, J. F. Hickey, J. M. Escaravage, L. Esau, A. T. Dore, S. Stevens, N. C. Adams, M. G. Dominguez, N. W. Gale, G. D. Yancopoulos, T. M. DeChiara, and D. M. Valenzuela. F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nat. Biotechnol.*, 25(1):91–99, Jan. 2007.

C. Pouget, T. Peterkin, F. C. Simoes, Y. Lee, D. Traver, and R. Patient. FGF signalling restricts haematopoietic stem cell specification via modulation of the BMP pathway. *Nat Commun*, 5:5588, 2014.

A. Pozner, D. Goldenberg, V. Negreanu, S. Y. Le, O. Elroy-Stein, D. Levanon, and Y. Groner. Transcription-coupled translation control of AML1/RUNX1 is mediated by cap- and internal ribosome entry site-dependent mechanisms. *Mol. Cell. Biol.*, 20(7):2297–2307, Apr. 2000.

M. L. Prato Murphy, M. A. Forsyth, G. J. Belsham, and J. S. Salt. Localization of foot-and-mouth disease virus RNA by in situ hybridization within bovine tissues. *Virus Res.*, 62(1): 67–76, July 1999.

C. Proudfoot, D. F. Carlson, R. Huddart, C. R. Long, J. H. Pryor, T. J. King, S. G. Lillico, A. J.

Mileham, D. G. McLaren, C. B. A. Whitelaw, and S. C. Fahrenkrug. Genome edited sheep and cattle. *Transgenic Res.*, Sept. 2014.

L. Ramirez-Carvajal, F. Diaz-San Segundo, D. Hickman, C. R. Long, J. Zhu, L. L. Rodriguez, and T. de los Santos. Expression of porcine fusion protein IRF7/3(5d) efficiently controls foot-and-mouth disease virus replication. *J. Virol.*, 88(19):11140–11153, Oct. 2014.

F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 8(11):2281–2308, Nov. 2013.

B. K. Ray, T. G. Lawson, J. C. Kramer, M. H. Cladaras, J. A. Grifo, R. D. Abramson, W. C. Merrick, and R. E. Thach. ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J. Biol. Chem.*, 260(12):7651–7658, June 1985.

A. Rebbapragada, H. Benchabane, J. L. Wrana, A. J. Celeste, and L. Attisano. Myostatin signals through a transforming growth factor beta-like signaling pathway to block adipogenesis. *Mol. Cell. Biol.*, 23(20):7230–7242, Oct. 2003.

N. Redondo, M. A. Sanz, E. Welnowska, and L. Carrasco. Translation without eIF2 promoted by poliovirus 2a protease. *PLoS ONE*, 6(10):e25699, 2011.

E. Riballo, L. Woodbine, T. Stiff, S. A. Walker, A. A. Goodarzi, and P. A. Jeggo. XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Res.*, 37(2):482–492, Feb. 2009.

J. A. Richt, P. Kasinathan, A. N. Hamir, J. Castilla, T. Sathiyaseelan, F. Vargas, J. Sathiyaseelan, H. Wu, H. Matsushita, J. Koster, S. Kato, I. Ishida, C. Soto, J. M. Robl, and Y. Kuroiwa. Production of cattle lacking prion protein. *Nat. Biotechnol.*, 25(1):132–138, Jan. 2007.

C. Richter, J. T. Chang, and P. C. Fineran. Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. *Viruses*, 4 (10):2291–2311, Oct. 2012.

R. Rios, S. Fernandez-Nocelos, I. Carneiro, V. M. Arce, and J. Devesa. Differential response to exogenous and endogenous myostatin in myoblasts suggests that myostatin acts as an autocrine factor in vivo. *Endocrinology*, 145(6):2795–2803, June 2004.

N. Rivera-Torres, B. Strouse, P. Bialk, R. A. Niamat, and E. B. Kmiec. The Position of DNA Cleavage by TALENs and Cell Synchronization Influences the Frequency of Gene Editing Directed by Single-Stranded Oligonucleotides. *PLoS ONE*, 9(5):e96483, 2014.

L. O. Roberts, R. A. Seamons, and G. J. Belsham. Recognition of picornavirus internal ribo-

some entry sites within cells; influence of cellular and viral proteins. *RNA*, 4(5):520–529, May 1998.

P. J. Roberts and G. J. Belsham. Identification of critical amino acids within the foot-and-mouth disease virus leader protein, a cysteine protease. *Virology*, 213(1):140–146, Oct. 1995.

J. Rodriguez, B. Vernus, M. Toubiana, E. Jublanc, L. Tintignac, S. Leibovitch, and A. Bonnieu. Myostatin inactivation increases myotube size through regulation of translational initiation machinery. *J. Cell. Biochem.*, 112(12):3531–3542, Dec. 2011.

M. Rodriguez Pulido, P. Serrano, M. Saiz, and E. MartÃnez-Salas. Foot-and-mouth disease virus infection induces proteolytic cleavage of PTB, eIF3a,b, and PABP RNA-binding proteins. *Virology*, 364(2):466–474, Aug. 2007.

E. P. Rogakou, D. R. Pilch, A. H. Orr, V. S. Ivanova, and W. M. Bonner. DNA double-stranded breaks induce histone H2ax phosphorylation on serine 139. *J. Biol. Chem.*, 273(10):5858–5868, Mar. 1998.

C. S. Rogers, Y. Hao, T. Rokhlina, M. Samuel, D. A. Stoltz, Y. Li, E. Petroff, D. W. Vermeer, A. C. Kabel, Z. Yan, L. Spate, D. Wax, C. N. Murphy, A. Rieke, K. Whitworth, M. L. Linville, S. W. Korte, J. F. Engelhardt, M. J. Welsh, and R. S. Prather. Production of CFTR-null and CFTR-DeltaF508 heterozygous pigs by adeno-associated virus-mediated gene targeting and somatic cell nuclear transfer. *J. Clin. Invest.*, 118(4):1571–1577, Apr. 2008a.

C. S. Rogers, D. A. Stoltz, D. K. Meyerholz, L. S. Ostedgaard, T. Rokhlina, P. J. Taft, M. P. Rogan, A. A. Pezzulo, P. H. Karp, O. A. Itani, A. C. Kabel, C. L. Wohlford-Lenane, G. J. Davis, R. A. Hanfland, T. L. Smith, M. Samuel, D. Wax, C. N. Murphy, A. Rieke, K. Whitworth, A. Uc, T. D. Starner, K. A. Brogden, J. Shilyansky, P. B. McCray, Jr, J. Zabner, R. S. Prather, and M. J. Welsh. Disruption of the CFTR gene produces a model of cystic fibrosis in newborn pigs. *Science*, 321(5897):1837–1841, Sept. 2008b.

K. Ross, A. K. Sedello, G. P. Todd, M. Paszkowski-Rogacz, A. W. Bird, L. Ding, T. Grinenko, K. Behrens, N. Hubner, M. Mann, C. Waskow, C. Stocking, and F. Buchholz. Polycomb group ring finger 1 cooperates with Runx1 in regulating differentiation and self-renewal of hematopoietic cells. *Blood*, 119(18):4152–4161, May 2012.

J. Rossant. Investigation of inner cell mass determination by aggregation of isolated rat inner cell masses with mouse morulae. *J Embryol Exp Morphol*, 36(1):163–174, Aug. 1976.

J. Rossant and W. I. Frels. Interspecific chimeras in mammals: successful production of live chimeras between Mus musculus and Mus caroli. *Science*, 208(4442):419–421, Apr. 1980.

D. B. Roth and J. H. Wilson. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. *Mol. Cell. Biol.*, 6(12):4295–4304, Dec. 1986.

D. B. Roth, T. N. Porter, and J. H. Wilson. Mechanisms of nonhomologous recombination in mammalian cells. *Mol. Cell. Biol.*, 5(10):2599–2607, Oct. 1985.

D. B. Roth, X. B. Chang, and J. H. Wilson. Comparison of filler DNA at immune, nonimmune, and oncogenic rearrangements suggests multiple mechanisms of formation. *Mol. Cell. Biol.*, 9(7):3049–3057, July 1989.

P. Rouet, F. Smih, and M. Jasin. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.*, 14(12):8096–8106, Dec. 1994.

G. M. Rubin and A. C. Spradling. Genetic transformation of Drosophila with transposable element vectors. *Science*, 218(4570):348–353, Oct. 1982.

J. Ruiz-Saenz, Y. Goez, W. Tabares, and A. Lopez-Herrera. Cellular receptors for foot and mouth disease virus. *Intervirology*, 52(4):201–212, 2009.

J. Rygaard and C. O. Povlsen. Heterotransplantation of a human malignant tumour to "Nude" mice. *Acta Pathol Microbiol Scand*, 77(4):758–760, 1969.

M. J. Sadofsky, J. E. Hesse, J. F. McBlane, and M. Gellert. Expression and V(D)J recombination activity of mutated RAG-1 proteins. *Nucleic Acids Res.*, 21(24):5644–5650, Dec. 1993.

E. Sakai, K. Kitajima, A. Sato, and T. Nakano. Increase of hematopoietic progenitor and suppression of endothelial gene expression by Runx1 expression during in vitro ES differentiation. *Exp. Hematol.*, 37(3):334–345, Mar. 2009.

T. Sakuma, H. Ochiai, T. Kaneko, T. Mashimo, D. Tokumasu, Y. Sakane, K.-i. Suzuki, T. Miyamoto, N. Sakamoto, S. Matsuura, and T. Yamamoto. Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity. *Sci Rep*, 3:3379, 2013.

I. M. Samokhvalov, A. M. Thomson, C. Lalancette, A. Liakhovitskaia, J. Ure, and A. Medvinsky. Multifunctional reversible knockout/reporter system enabling fully functional recon-

stitution of the AML1/Runx1 locus and rescue of hematopoiesis. *Genesis*, 44(3):115–121, Mar. 2006.

S. Sarrazin and M. Sieweke. Integration of cytokine and transcription factor signals in hematopoietic stem cell commitment. *Semin. Immunol.*, 23(5):326–334, Oct. 2011.

K. Sasaki, H. Yagi, R. T. Bronson, K. Tominaga, T. Matsunashi, K. Deguchi, Y. Tani, T. Kishimoto, and T. Komori. Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor beta. *Proc. Natl. Acad. Sci. U.S.A.*, 93(22): 12359–12363, Oct. 1996.

D. G. Schatz and Y. Ji. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.*, 11(4):251–263, Apr. 2011.

D. G. Schatz, M. A. Oettinger, and D. Baltimore. The V(D)J recombination activating gene, RAG-1. *Cell*, 59(6):1035–1048, Dec. 1989.

R. Schofield. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells*, 4(1-2):7–25, 1978.

J. Seipelt, A. GuarnÃ©, E. Bergmann, M. James, W. Sommergruber, I. Fita, and T. Skern. The structures of picornaviral proteinases. *Virus Res.*, 62(2):159–168, Aug. 1999.

L. M. Seligman, K. M. Chisholm, B. S. Chevalier, M. S. Chadsey, S. T. Edwards, J. H. Savage, and A. L. Veillet. Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.*, 30(17):3870–3879, Sept. 2002.

O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, and F. Zhang. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, 343(6166):84–87, Jan. 2014.

N. M. Shanbhag, I. U. Rafalska-Metcalf, C. Balane-Bolivar, S. M. Janicki, and R. A. Greenberg. ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell*, 141(6):970–981, June 2010.

D. Shaw, W. Dondorp, N. Geijsen, and G. de Wert. Creating human organs in chimaera pigs: an ethical source of immunocompatible organs? *J Med Ethics*, Nov. 2014.

H. Shi, Q. Fu, Y. Ren, D. Wang, J. Qiao, P. Wang, H. Zhang, and C. Chen. Both foot-and-mouth disease virus and bovine viral diarrhea virus replication are inhibited by Mx1 protein originated from porcine. *Anim. Biotechnol.*, 26(1):73–79, 2015.

H. Shima, H. Suzuki, J. Sun, K. Kono, L. Shi, A. Kinomura, Y. Horikoshi, T. Ikura, M. Ikura,

R. Kanaar, K. Igarashi, H. Saitoh, H. Kurumizaka, and S. Tashiro. Activation of the SUMO modification system is required for the accumulation of RAD51 at sites of DNA damage. *J. Cell. Sci.*, 126(Pt 22):5284–5292, Nov. 2013.

G. H. Silva, M. Belfort, W. Wende, and A. Pingoud. From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. *J. Mol. Biol.*, 361(4):744–754, Aug. 2006.

D. P. Silver, E. Spanopoulou, R. C. Mulligan, and D. Baltimore. Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. *Proc. Natl. Acad. Sci. U.S.A.*, 90(13):6100–6104, July 1993.

C. Simerly, D. McFarland, C. Castro, C.-C. Lin, C. Redinger, E. Jacoby, J. Mich-Basso, K. Orwig, P. Mills, E. Ahrens, C. Navara, and G. Schatten. Interspecies chimera between primate embryonic stem cells and mouse embryos: monkey ESCs engraft into mouse embryos, but not post-implantation fetuses. *Stem Cell Res*, 7(1):28–40, July 2011.

D. Simsek and M. Jasin. Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat. Struct. Mol. Biol.*, 17 (4):410–416, Apr. 2010.

D. Simsek, E. Brunet, S. Y.-W. Wong, S. Katyal, Y. Gao, P. J. McKinnon, J. Lou, L. Zhang, J. Li, E. J. Rebar, P. D. Gregory, M. C. Holmes, and M. Jasin. DNA ligase III promotes alternative nonhomologous end-joining during chromosomal translocation formation. *PLoS Genet.*, 7(6):e1002080, June 2011.

C. Smith, A. Gore, W. Yan, L. Abalde-Atristain, Z. Li, C. He, Y. Wang, R. A. Brodsky, K. Zhang, L. Cheng, and Z. Ye. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell*, 15(1): 12–13, July 2014.

J. Smith, J. M. Berg, and S. Chandrasegaran. A detailed study of the substrate specificity of a chimeric restriction enzyme. *Nucleic Acids Res.*, 27(2):674–681, Jan. 1999.

J. Smith, M. Bibikova, F. G. Whitby, A. R. Reddy, S. Chandrasegaran, and D. Carroll. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.*, 28(17):3361–3369, Sept. 2000.

J. Smith, S. Grizot, S. Arnould, A. Duclert, J.-C. Epinat, P. Chames, J. Prieto, P. Redondo, F. J. Blanco, J. Bravo, G. Montoya, F. Paques, and P. Duchateau. A combinatorial approach to

create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.*, 34 (22):e149, 2006.

J. Sorge and S. H. Hughes. Splicing of intervening sequences introduced into an infectious retroviral vector. *J. Mol. Appl. Genet.*, 1(6):547–559, 1982.

D. Y. Soung, L. Talebian, C. J. Matheny, R. Guzzo, M. E. Speck, J. R. Lieberman, N. A. Speck, and H. Drissi. Runx1 dose-dependently regulates endochondral ossification during skeletal development and fracture healing. *J. Bone Miner. Res.*, 27(7):1585–1597, July 2012.

P. Sroczynska, C. Lancrin, V. Kouskoff, and G. Lacaud. The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis. *Blood*, 114(26):5279–5289, Dec. 2009.

P. B. Staber, P. Zhang, M. Ye, R. S. Welner, E. Levantini, A. Di Ruscio, A. K. Ebralidze, C. Bach, H. Zhang, J. Zhang, K. Vanura, R. Delwel, H. Yang, G. Huang, and D. G. Tenen. The Runx-PU.1 pathway preserves normal and AML/ETO9a leukemic stem cells. *Blood*, 124(15):2391–2399, Oct. 2014.

P. Stanhope-Baker, K. M. Hudson, A. L. Shaffer, A. Constantinescu, and M. S. Schlissel. Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell*, 85(6):887–897, June 1996.

C. Stenfeldt, J. M. Pacheco, L. L. Rodriguez, and J. Arzt. Early events in the pathogenesis of foot-and-mouth disease in pigs; identification of oropharyngeal tonsils as sites of primary and sustained viral replication. *PLoS ONE*, 9(9):e106859, 2014a.

C. Stenfeldt, J. M. Pacheco, G. R. Smoliga, E. Bishop, S. J. Pauszek, E. J. Hartwig, L. L. Rodriguez, and J. Arzt. Detection of Foot-and-mouth Disease Virus RNA and Capsid Protein in Lymphoid Tissues of Convalescent Pigs Does Not Indicate Existence of a Carrier State. *Transbound Emerg Dis*, June 2014b.

M. S. Stern. Letter: Chimaeras obtained by aggregation of mouse eggs with rat eggs. *Nature*, 243(5408):472–473, June 1973.

S. Steuer, V. Pingoud, A. Pingoud, and W. Wende. Chimeras of the homing endonuclease PI-SceI and the homologous Candida tropicalis intein: a study to explore the possibility of exchanging DNA-binding modules to obtain highly specific endonucleases with altered specificity. *Chembiochem*, 5(2):206–213, Feb. 2004.

N. Stifani, A. R. O. Freitas, A. Liakhovitskaia, A. Medvinsky, A. Kania, and S. Stifani. Sup-

pression of interneuron programs and maintenance of selected spinal motor neuron fates by the transcription factor AML1/Runx1. *Proc. Natl. Acad. Sci. U.S.A.*, 105(17):6451–6456, Apr. 2008.

E. Stillwell, J. Vitale, Q. Zhao, A. Beck, J. Schneider, F. Khadim, G. Elson, A. Altaf, G. Yehia, J.-h. Dong, J. Liu, W. Mark, M. Bhaumik, R. Grange, and D. Fraidenraich. Blastocyst injection of wild type embryonic stem cells induces global corrections in mdx mice. *PLoS ONE*, 4(3):e4759, 2009.

D. T. Stinchcomb, J. E. Shaw, S. H. Carr, and D. Hirsh. Extrachromosomal DNA transformation of Caenorhabditis elegans. *Mol. Cell. Biol.*, 5(12):3484–3496, Dec. 1985.

B. L. Stoddard. Homing endonuclease structure and function. *Q. Rev. Biophys.*, 38(1):49–95, Feb. 2005.

N. Strande, S. A. Roberts, S. Oh, E. A. Hendrickson, and D. A. Ramsden. Specificity of the dRP/AP lyase of Ku promotes nonhomologous end joining (NHEJ) fidelity at damaged ends. *J. Biol. Chem.*, 287(17):13686–13693, Apr. 2012.

K. Strebel and E. Beck. A second protease of foot-and-mouth disease virus. *J. Virol.*, 58(3): 893–899, June 1986.

R. Strong and G. J. Belsham. Sequential modification of translation initiation factor eIF4gi by two different foot-and-mouth disease virus proteases within infected baby hamster kidney cells: identification of the 3cpro cleavage site. *J. Gen. Virol.*, 85(Pt 10):2953–2962, Oct. 2004.

R. Subrahmanyam and R. Sen. RAGs' eye view of the immunoglobulin heavy chain gene locus. *Semin. Immunol.*, 22(6):337–345, Dec. 2010.

Y. H. Sung, I.-J. Baek, D. H. Kim, J. Jeon, J. Lee, K. Lee, D. Jeong, J.-S. Kim, and H.-W. Lee. Knockout mice created by TALEN-mediated gene targeting. *Nat. Biotechnol.*, 31(1):23–24, Jan. 2013.

D. Sussman, M. Chadsey, S. Fauce, A. Engel, A. Bruett, R. Monnat, Jr, B. L. Stoddard, and L. M. Seligman. Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.*, 342(1):31–41, Sept. 2004.

K. Suzuki, C. Yu, J. Qu, M. Li, X. Yao, T. Yuan, A. Goebl, S. Tang, R. Ren, E. Aizawa, F. Zhang, X. Xu, R. Soligalla, F. Chen, J. Kim, N. Kim, H.-K. Liao, C. Benner, C. Esteban, Y. Jin, G.-H. Liu, Y. Li, and J. IzpisuaÂ Belmonte. Targeted Gene Correction Minimally

Impacts Whole-Genome Mutational Load in Human-Disease-Specific Induced Pluripotent Stem Cell Clones. *Cell Stem Cell*, 15(1):31–36, July 2014.

J. M. Svendsen and J. W. Harper. GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution. *Genes Dev.*, 24(6):521–536, Mar. 2010.

J. M. Svendsen, A. Smogorzewska, M. E. Sowa, B. C. O'Connell, S. P. Gygi, S. J. Elledge, and J. W. Harper. Mammalian BTBD12/SLX4 Assembles A Holliday Junction Resolvase and Is Required for DNA Repair. *Cell*, 138(1):63–77, July 2009.

P. C. Swanson. A RAG-1/RAG-2 tetramer supports 12/23-regulated synapsis, cleavage, and transposition of V(D)J recombination signals. *Mol. Cell. Biol.*, 22(22):7790–7801, Nov. 2002.

G. Szabo, G. Dallmann, G. Mueller, L. Patthy, M. Soller, and L. Varga. A deletion in the myostatin gene causes the compact (Cmpt) hypermuscular mutation in mice. *Mamm. Genome*, 9(8):671–672, Aug. 1998.

M. Szczepek, V. Brondani, J. BÃŒchel, L. Serrano, D. J. Segal, and T. Cathomen. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.*, 25(7):786–793, July 2007.

E. H. Szybalska and W. Szybalski. Genetics of human cess line. IV. DNA-mediated heritable transformation of a biochemical trait. *Proc. Natl. Acad. Sci. U.S.A.*, 48:2026–2034, Dec. 1962.

M. Tachibana, M. Sparman, C. Ramsey, H. Ma, H.-S. Lee, M. C. T. Penedo, and S. Mitalipov. Generation of chimeric rhesus monkeys. *Cell*, 148(1-2):285–295, Jan. 2012.

S. Takada, T. Sato, Y. Ito, S. Yamashita, T. Kato, M. Kawasumi, M. Kanai-Azuma, A. Igarashi, T. Kato, M. Tamano, and H. Asahara. Targeted Gene Deletion of miRNAs in Mice by TALEN System. *PLoS ONE*, 8(10):e76004, 2013.

K. Takahashi and S. Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–676, Aug. 2006.

H. Takeda, C. Charlier, F. Farnir, and M. Georges. Demonstrating polymorphic miRNA-mediated gene regulation in vivo: application to the g+6223g->A mutation of Texel sheep. *RNA*, 16(9):1854–1863, Sept. 2010.

W. Tan, D. F. Carlson, C. A. Lancto, J. R. Garbe, D. A. Webster, P. B. Hackett, and S. C.

Fahrenkrug. Efficient nonmeiotic allele introgression in livestock using custom endonucleases. *Proc. Natl. Acad. Sci. U.S.A.*, Sept. 2013.

Y. Tanaka, A. Joshi, N. K. Wilson, S. Kinston, S. Nishikawa, and B. Goettgens. The transcriptional programme controlled by Runx1 during early embryonic blood development. *Developmental Biology*, Apr. 2012.

I. Taniuchi, M. Osato, T. Egawa, M. J. Sunshine, S. C. Bae, T. Komori, Y. Ito, and D. R. Littman. Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. *Cell*, 111(5):621–633, Nov. 2002.

S. Taoudi, C. Gonneau, K. Moore, J. M. Sheridan, C. C. Blackburn, E. Taylor, and A. Medvinsky. Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-cadherin+CD45+ pre-definitive HSCs. *Cell Stem Cell*, 3(1):99–108, July 2008.

A. K. Tarkowski. Mouse chimaeras developed from fused eggs. *Nature*, 190:857–860, June 1961.

K. Tashiro, K. Kawabata, M. Omori, T. Yamaguchi, F. Sakurai, K. Katayama, T. Hayakawa, and H. Mizuguchi. Promotion of hematopoietic differentiation from mouse induced pluripotent stem cells by transient HoxB4 transduction. *Stem Cell Res*, 8(2):300–311, Mar. 2012.

E. F. Terwilliger, B. Godin, J. G. Sodroski, and W. A. Haseltine. Construction and use of a replication-competent human immunodeficiency virus (HIV-1) that expresses the chloramphenicol acetyltransferase enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, 86(10):3857–3861, May 1989.

P. J. Tesar, J. G. Chenoweth, F. A. Brook, T. J. Davies, E. P. Evans, D. L. Mack, R. L. Gardner, and R. D. G. McKay. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, 448(7150):196–199, July 2007.

K. R. Thomas and M. R. Capecchi. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, 51(3):503–512, Nov. 1987.

G. R. Thomson, W. Vosloo, and A. D. S. Bastos. Foot and mouth disease in wildlife. *Virus Res.*, 91(1):145–161, Jan. 2003.

J. A. Thomson, J. Kalishman, T. G. Golos, M. Durning, C. P. Harris, R. A. Becker, and J. P. Hearn. Isolation of a primate embryonic stem cell line. *Proc. Natl. Acad. Sci. U.S.A.*, 92(17):7844–7848, Aug. 1995.

J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall,

and J. M. Jones. Embryonic stem cell lines derived from human blastocysts. *Science*, 282 (5391):1145–1147, Nov. 1998.

M. R. Tijssen, A. Cvejic, A. Joshi, R. L. Hannah, R. Ferreira, A. Forrai, D. C. Bellissimo, S. H. Oram, P. A. Smethurst, N. K. Wilson, X. Wang, K. Ottersbach, D. L. Stemple, A. R. Green, W. H. Ouwehand, and B. Goettgens. Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev. Cell*, 20(5):597–609, May 2011.

J. E. Till and E. A. McCulloch. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat. Res.*, 14:213–222, Feb. 1961.

C. Tong, P. Li, N. L. Wu, Y. Yan, and Q.-L. Ying. Production of p53 gene knockout rats by homologous recombination in embryonic stem cells. *Nature*, 467(7312):211–213, Sept. 2010.

A. U. Trendelenburg, A. Meyer, D. Rohner, J. Boyle, S. Hatakeyama, and D. J. Glass. Myostatin reduces Akt/TORC1/p70s6k signaling, inhibiting myoblast differentiation and myotube size. *Am. J. Physiol., Cell Physiol.*, 296(6):C1258–1270, June 2009.

S. Q. Tsai, N. Wyvekens, C. Khayter, J. A. Foden, V. Thapar, D. Reyon, M. J. Goodwin, M. J. Aryee, and J. K. Joung. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.*, 32(6):569–576, June 2014.

E. M. Tucker, R. M. Moor, and L. E. Rowson. Tetraparental sheep chimaeras induced by blastomere transplantation. Changes in blood type with age. *Immunology*, 26(3):613–621, Mar. 1974.

N. Uematsu, E. Weterings, K.-i. Yano, K. Morotomi-Yano, B. Jakob, G. Taucher-Scholz, P.-O. Mari, D. C. van Gent, B. P. C. Chen, and D. J. Chen. Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J. Cell Biol.*, 177(2):219–229, Apr. 2007.

F. D. Urnov, J. C. Miller, Y.-L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory, and M. C. Holmes. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435(7042):646–651, June 2005.

J.-I. Usui, T. Kobayashi, T. Yamaguchi, A. S. Knisely, R. Nishinakamura, and H. Nakauchi. Generation of Kidney from Pluripotent Stem Cells via Blastocyst Complementation. *The American Journal of Pathology*, Apr. 2012.

G. Van den Ackerveken, E. Marois, and U. Bonas. Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell*, 87(7):1307–1316, Dec. 1996.

B. van Riel, T. Pakozdi, R. Brouwer, R. Monteiro, K. Tuladhar, V. Franke, J. C. Bryne, R. Jorna, E.-J. Rijkers, W. van Ijcken, C. Andrieu-Soler, J. Demmers, R. Patient, E. Soler, B. Lenhard, and F. Grosveld. A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells. *Mol. Cell. Biol.*, 32(19):3814–3822, Oct. 2012.

P. Van Roey, C. A. Waddling, K. M. Fox, M. Belfort, and V. Derbyshire. Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J.*, 20 (14):3631–3637, July 2001.

E. S. Vanamee, S. Santagata, and A. K. Aggarwal. FokI requires two specific DNA sites for cleavage. *J. Mol. Biol.*, 309(1):69–78, May 2001.

L. Varga, G. Szabo, A. Darvasi, G. Mueller, M. Sass, and M. Soller. Inheritance and mapping of Compact (Cmpt), a new mutation causing hypermuscularity in mice. *Genetics*, 147(2): 755–764, Oct. 1997.

A. Veres, B. S. Gosis, Q. Ding, R. Collins, A. Ragavendran, H. Brand, S. Erdin, M. E. Talkowski, and K. Musunuru. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell*, 15(1):27–30, July 2014.

L. Vouillot, A. Thelie, and N. Pollet. Comparison of T7e1 and Surveyor Mismatch Cleavage Assays To Detect Mutations Triggered by Engineered Nucleases. *G3 (Bethesda)*, Jan. 2015.

E. J. Walker, P. Younessi, A. J. Fulcher, R. McCuaig, B. J. Thomas, P. G. Bardin, D. A. Jans, and R. Ghildyal. Rhinovirus 3c protease facilitates specific nucleoporin cleavage and mislocalisation of nuclear proteins in infected host cells. *PLoS ONE*, 8(8):e71316, 2013.

J. R. Walker, R. A. Corpina, and J. Goldberg. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*, 412(6847):607–614, Aug. 2001.

R. J. Wall, A. M. Powell, M. J. Paape, D. E. Kerr, D. D. Bannerman, V. G. Pursel, K. D. Wells, N. Talbot, and H. W. Hawk. Genetically enhanced cows resist intramammary Staphylococcus aureus infection. *Nat. Biotechnol.*, 23(4):445–451, Apr. 2005.

C. Q. Wang, L. Motoda, M. Satake, Y. Ito, I. Taniuchi, V. Tergaonkar, and M. Osato. Runx3 deficiency results in myeloproliferative disorder in aged mice. *Blood*, 122(4):562–566, July 2013a.

D. Wang, L. Fang, R. Luo, R. Ye, Y. Fang, L. Xie, H. Chen, and S. Xiao. Foot-and-mouth disease virus leader proteinase inhibits dsRNA-induced type I interferon transcription by decreasing interferon regulatory factor 3/7 in protein levels. *Biochem. Biophys. Res. Commun.*, 399(1):72–78, Aug. 2010.

D. Wang, L. Fang, J. Bi, Q. Chen, L. Cao, R. Luo, H. Chen, and S. Xiao. Foot-and-mouth disease virus leader proteinase inhibits dsRNA-induced RANTES transcription in PK-15 cells. *Virus Genes*, 42(3):388–393, June 2011a.

D. Wang, L. Fang, P. Li, L. Sun, J. Fan, Q. Zhang, R. Luo, X. Liu, K. Li, H. Chen, Z. Chen, and S. Xiao. The leader proteinase of foot-and-mouth disease virus negatively regulates the type I interferon pathway by acting as a viral deubiquitinase. *J. Virol.*, 85(8):3758–3766, Apr. 2011b.

D. Wang, L. Fang, K. Li, H. Zhong, J. Fan, C. Ouyang, H. Zhang, E. Duan, R. Luo, Z. Zhang, X. Liu, H. Chen, and S. Xiao. Foot-and-mouth disease virus 3c protease cleaves NEMO to impair innate immune signaling. *J. Virol.*, 86(17):9311–9322, Sept. 2012a.

D. Wang, L. Fang, D. Wei, H. Zhang, R. Luo, H. Chen, K. Li, and S. Xiao. Hepatitis A virus 3c protease cleaves NEMO to impair induction of beta interferon. *J. Virol.*, 88(17): 10252–10258, Sept. 2014a.

H. Wang and S. K. Dey. Roadmap to embryo implantation: clues from mouse models. *Nature Reviews Genetics*, 7(3):185–199, Mar. 2006.

H. Wang, X. Liu, J. Wu, G. Wu, L. Yu, C. He, H. Yang, W. Xie, X. Xia, and H. He. Bovine fetal epithelium cells expressing shRNA targeting viral VP1 gene resisted against foot-and-mouth disease virus. *Virology*, 439(2):115–121, May 2013b.

L. D. Wang and A. J. Wagers. Dynamic niches in the origination and differentiation of haematopoietic stem cells. *Nat. Rev. Mol. Cell Biol.*, 12(10):643–655, Oct. 2011.

Q. Wang, T. Stacy, M. Binder, M. Marin-Padilla, A. H. Sharpe, and N. A. Speck. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl. Acad. Sci. U.S.A.*, 93(8):3444–3449, Apr. 1996.

S. Wang, J. Bates, X. Li, S. Schanz, D. Chandler-Militello, C. Levine, N. Maherali, L. Studer, K. Hochedlinger, M. Windrem, and S. A. Goldman. Human iPSC-derived oligodendrocyte progenitor cells can myelinate and rescue a mouse model of congenital hypomyelination. *Cell Stem Cell*, 12(2):252–264, Feb. 2013c.

T. Wang, J. J. Wei, D. M. Sabatini, and E. S. Lander. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, 343(6166):80–84, Jan. 2014b.

X. Wang, E. Montini, M. Al-Dhalimy, E. Lagasse, M. Finegold, and M. Grompe. Kinetics of liver repopulation after bone marrow transplantation. *Am. J. Pathol.*, 161(2):565–574, Aug. 2002.

X. Wang, C. Blagden, J. Fan, S. J. Nowak, I. Taniuchi, D. R. Littman, and S. J. Burden. Runx1 prevents wasting, myofibrillar disorganization, and autophagy of skeletal muscle. *Genes Dev.*, 19(14):1715–1722, July 2005.

X. Wang, G. Xiao, Y. Zhang, X. Wen, X. Gao, S. Okada, and X. Liu. Regulation of Tcrb recombination ordering by c-Fos-dependent RAG deposition. *Nat. Immunol.*, 9(7):794–801, July 2008.

X. Wang, Y. Wang, X. Wu, J. Wang, Y. Wang, Z. Qiu, T. Chang, H. Huang, R.-J. Lin, and J.-K. Yee. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, Jan. 2015.

Y. Wang, X.-Y. Zhou, P.-Y. Xiang, L.-L. Wang, H. Tang, F. Xie, L. Li, and H. Wei. The Meganuclease I-SceI Containing Nuclear Localization Signal (NLS-I-SceI) Efficiently Mediated Mammalian Germline Transgenesis via Embryo Cytoplasmic Microinjection. *PLoS ONE*, 9 (9):e108347, 2014c.

Y. X. Wang, N. A. Dumont, and M. A. Rudnicki. Muscle stem cells at a glance. *J. Cell. Sci.*, Oct. 2014d.

Z. Wang, E. Oron, B. Nelson, S. Razis, and N. Ivanova. Distinct Lineage Specification Roles for NANOG, OCT4, and SOX2 in Human Embryonic Stem Cells. *Cell Stem Cell*, 10(4): 440–454, Apr. 2012b.

M. Watanabe, K. Umeyama, H. Matsunari, S. Takayanagi, E. Haruyama, K. Nakano, T. Fujiwara, Y. Ikezawa, H. Nakauchi, and H. Nagashima. Knockout of exogenous EGFP gene in porcine somatic cells using zinc-finger nucleases. *Biochem. Biophys. Res. Commun.*, 402(1): 14–18, Nov. 2010.

G. V. Weaver, J. Domenech, A. R. Thiermann, and W. B. Karesh. Foot and mouth disease: a look from the wild side. *J. Wildl. Dis.*, 49(4):759–785, Oct. 2013.

E. M. Weber, B. Algers, J. Hultgren, and I. A. S. Olsson. Pup mortality in laboratory mice infanticide or not? *Acta Veterinaria Scandinavica*, 55(1):83, 2013.

J. Wei, H. Ouyang, Y. Wang, D. Pang, N. X. Cong, T. Wang, B. Leng, D. Li, X. Li, R. Wu, Y. Ding, F. Gao, Y. Deng, B. Liu, Z. Li, L. Lai, H. Feng, G. Liu, and X. Deng. Characterization of a hypertriglyceridemic transgenic miniature pig model expressing human apolipoprotein CIII. *FEBS J.*, 279(1):91–99, Jan. 2012.

S. Welle, K. Bhatt, C. A. Pinkert, R. Tawil, and C. A. Thornton. Muscle growth after postdevelopmental myostatin gene knockout. *Am. J. Physiol. Endocrinol. Metab.*, 292(4):E985–991, Apr. 2007.

E. Welnowska, M. A. Sanz, N. Redondo, and L. Carrasco. Translation of viral mRNA without active eIF2: the case of picornaviruses. *PLoS ONE*, 6(7):e22230, 2011.

K. L. West, N. C. Singha, P. De Ioannes, L. Lacomis, H. Erdjument-Bromage, P. Tempst, and P. Cortes. A direct interaction between the RAG2 C terminus and the core histones is required for efficient V(D)J recombination. *Immunity*, 23(2):203–212, Aug. 2005.

S. C. West. The search for a human Holliday junction resolvase. *Biochem. Soc. Trans.*, 37(Pt 3):519–526, June 2009.

L.-A. Whittemore, K. Song, X. Li, J. Aghajanian, M. Davies, S. Girgenrath, J. J. Hill, M. Jalenak, P. Kelley, A. Knight, R. Maylor, D. O'Hara, A. Pearson, A. Quazi, S. Ryerson, X. Y. Tan, K. N. Tomkinson, G. M. Veldman, A. Widom, J. F. Wright, S. Wudyka, L. Zhao, and N. M. Wolfman. Inhibition of myostatin in adult mice increases skeletal muscle mass and strength. *Biochem. Biophys. Res. Commun.*, 300(4):965–971, Jan. 2003.

WHO. *Global status report on noncommunicable diseases: 2010*. World Health Organization, Geneva, 2011. ISBN 9789241564229 9241564229.

I. Wilmut, A. E. Schnieke, J. McWhir, A. J. Kind, and K. H. Campbell. Viable offspring derived from fetal and adult mammalian cells. *Nature*, 385(6619):810–813, Feb. 1997.

E. M. Wilson, J. Bial, B. Tarlow, G. Bial, B. Jensen, D. L. Greiner, M. A. Brehm, and M. Grompe. Extensive double humanization of both liver and hematopoiesis in FRGN mice. *Stem Cell Res*, 13(3PA):404–412, Sept. 2014.

J. H. Wilson, P. B. Berget, and J. M. Pipas. Somatic cells efficiently join unrelated DNA segments end-to-end. *Mol. Cell. Biol.*, 2(10):1258–1269, Oct. 1982.

N. K. Wilson, S. D. Foster, X. Wang, K. Knezevic, J. Schuette, P. Kaimakis, P. M. Chilarska, S. Kinston, W. H. Ouwehand, E. Dzierzak, J. E. Pimanda, M. F. T. R. de Bruijn, and B. Goettgens. Combinatorial transcriptional control in blood stem/progenitor cells: genome-

wide analysis of ten major transcriptional regulators. *Cell Stem Cell*, 7(4):532–544, Oct. 2010.

E. Wolf, C. Braun-Reichhart, E. Streckel, and S. Renner. Genetically engineered pig models for diabetes research. *Transgenic Res.*, 23(1):27–38, Feb. 2014.

N. M. Wolfman, A. C. McPherron, W. N. Pappano, M. V. Davies, K. Song, K. N. Tomkinson, J. F. Wright, L. Zhao, S. M. Sebald, D. S. Greenspan, and S.-J. Lee. Activation of latent myostatin by the BMP-1/tolloid family of metalloproteinases. *Proc. Natl. Acad. Sci. U.S.A.*, 100(26):15842–15846, Dec. 2003.

G. Wright, A. Carver, D. Cottom, D. Reeves, A. Scott, P. Simons, I. Wilmut, I. Garner, and A. Colman. High level expression of active human alpha-1-antitrypsin in the milk of transgenic sheep. *Biotechnology (N.Y.)*, 9(9):830–834, Sept. 1991.

A. Xie, A. Kwok, and R. Scully. Role of mammalian Mre11 in classical and alternative non-homologous end joining. *Nat. Struct. Mol. Biol.*, 16(8):814–818, Aug. 2009.

J. Xu, P. Qian, Q. Wu, S. Liu, W. Fan, K. Zhang, R. Wang, H. Zhang, H. Chen, and X. Li. Swine interferon-induced transmembrane protein, sIFITM3, inhibits foot-and-mouth disease virus infection in vitro and in vivo. *Antiviral Res.*, 109:22–29, Sept. 2014.

X. Xu, J. Forrer, P. J. Bechtel, and P. M. Best. Elevated growth hormone increases the Ca2+ sensitivity of slow- and fast-twitch skeletal muscle of female rats. *Am. J. Physiol.*, 274(4 Pt 1):C861–865, Apr. 1998.

W. Xue, S. Chen, H. Yin, T. Tammela, T. Papagiannakopoulos, N. S. Joshi, W. Cai, G. Yang, R. Bronson, D. G. Crowley, F. Zhang, D. G. Anderson, P. A. Sharp, and T. Jacks. CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature*, 514(7522):380–384, Oct. 2014.

T. Yamashita and G. S. Sohal. Embryonic origin of skeletal muscle cells in the iris of the duck and quail. *Cell Tissue Res.*, 249(1):31–37, July 1987.

D. Yang, C.-E. Wang, B. Zhao, W. Li, Z. Ouyang, Z. Liu, H. Yang, P. Fan, A. O'Neill, W. Gu, H. Yi, S. Li, L. Lai, and X.-J. Li. Expression of Huntington's disease protein results in apoptotic neurons in the brains of cloned transgenic pigs. *Hum. Mol. Genet.*, 19(20):3983–3994, Oct. 2010.

H. Yang, H. Wang, C. S. Shivalila, A. W. Cheng, L. Shi, and R. Jaenisch. One-Step Genera-

tion of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*, Aug. 2013.

L. Yang, D. Grishin, G. Wang, J. Aach, C.-Z. Zhang, R. Chari, J. Homsy, X. Cai, Y. Zhao, J.-B. Fan, C. Seidman, J. Seidman, W. Pu, and G. Church. Targeted and genome-wide sequencing reveal single nucleotide variations impacting specificity of Cas9 in human stem cells. *Nat Commun*, 5:5507, 2014.

Y.-G. Yang and M. Sykes. Xenotransplantation: current status and a perspective on the future. *Nat. Rev. Immunol.*, 7(7):519–531, July 2007.

F. F. Yin, S. Bailey, C. A. Innis, M. Ciubotaru, S. Kamtekar, T. A. Steitz, and D. G. Schatz. Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat. Struct. Mol. Biol.*, 16(5):499–508, May 2009.

H. Yin, W. Xue, S. Chen, R. L. Bogorad, E. Benedetti, M. Grompe, V. Koteliansky, P. A. Sharp, T. Jacks, and D. G. Anderson. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.*, 32(6):551–553, June 2014.

J. J. Young, J. M. Cherone, Y. Doyon, I. Ankoudinova, F. M. Faraji, A. H. Lee, C. Ngo, D. Y. Guschin, D. E. Paschon, J. C. Miller, L. Zhang, E. J. Rebar, P. D. Gregory, F. D. Urnov, R. M. Harland, and B. Zeitler. Efficient targeted gene disruption in the soma and germ line of the frog Xenopus tropicalis using engineered zinc-finger nucleases. *Proceedings of the National Academy of Sciences*, 108(17):7052–7057, Apr. 2011.

C. Zhang, L. Wang, G. Ren, Z. Li, C. Ren, T. Zhang, K. Xu, and Z. Zhang. Targeted disruption of the sheep MSTN gene by engineered zinc-finger nucleases. *Mol. Biol. Rep.*, Nov. 2013.

Y. Zhang and M. Jasin. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. *Nat. Struct. Mol. Biol.*, 18(1):80–84, Jan. 2011.

Y. Zhang, E. Y. Shim, M. Davis, and S. E. Lee. Regulation of repair choice: Cdk1 suppresses recruitment of end joining factors at DNA breaks. *DNA Repair (Amst.)*, 8(10):1235–1241, Oct. 2009.

Z. Zhang, C. R. Espinoza, Z. Yu, R. Stephan, T. He, G. S. Williams, P. D. Burrows, J. Hagman, A. J. Feeney, and M. D. Cooper. Transcription factor Pax5 (BSAP) transactivates the RAG-mediated V(H)-to-DJ(H) rearrangement of immunoglobulin genes. *Nat. Immunol.*, 7(6): 616–624, June 2006.

B. Zhong, Y. Zhang, Y. Yan, Z. Wang, S. Ying, M. Huang, and F. Wang. MicroRNA-Mediated Myostatin Silencing in Caprine Fetal Fibroblasts. *PLoS ONE*, 9(9):e107071, 2014.

C. Zhu, K. D. Mills, D. O. Ferguson, C. Lee, J. Manis, J. Fleming, Y. Gao, C. C. Morton, and F. W. Alt. Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations. *Cell*, 109(7):811–821, June 2002.

X. Zhu, J. E. Yeadon, and S. J. Burden. AML1 is expressed in skeletal muscle and is regulated by innervation. *Mol. Cell. Biol.*, 14(12):8051–8057, Dec. 1994.

X. Zhu, M. Hadhazy, M. Wehling, J. G. Tidball, and E. M. McNally. Dominant negative myostatin produces hypertrophy without hyperplasia in muscle. *FEBS Lett.*, 474(1):71–75, May 2000.

T. A. Zimmers, M. V. Davies, L. G. Koniaris, P. Haynes, A. F. Esquela, K. N. Tomkinson, A. C. McPherron, N. M. Wolfman, and S.-J. Lee. Induction of cachexia in mice by systemically administered myostatin. *Science*, 296(5572):1486–1488, May 2002.

Y. Ziv, D. Bielopolski, Y. Galanty, C. Lukas, Y. Taya, D. C. Schultz, J. Lukas, S. Bekker-Jensen, J. Bartek, and Y. Shiloh. Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nat. Cell Biol.*, 8(8):870–876, Aug. 2006.

P. A. Zunszain, S. R. Knox, T. R. Sweeney, J. Yang, N. RoquÃ©-Rosell, G. J. Belsham, R. J. Leatherbarrow, and S. Curry. Insights into cleavage specificity from the crystal structure of foot-and-mouth disease virus 3c protease complexed with a peptide substrate. *J. Mol. Biol.*, 395(2):375–389, Jan. 2010.

A. Appendices

A.1. PCR primers and oligos

Primer name	Primer sequence (5'-3')	Annealing temperature	Purpose	Notes
Runx1 common	GAGTCCCAGCTGTCAATTCC	68°C	Genotyping of mice from Runx1 colony	Genotyping protocol from JAX mice website
Runx1 wildtype	GGTGATGGTCAGAGTGAAGC			
Runx1 mutant	TCGCAGCGCATCGCCTTCTA			
Rex1 forward	TTCTTGCCAGGTTCTGGAAGC	61°C	Distinguishes between mouse and rat DNA in chimeras	
Rex1 reverse	TTTCCCACACTCTGCACACAC			
Runx1 forward	AGATGCCAGCACGAGCCGCC	53°C	Amplifies TALEN target region in porcine Runx1	PCR reactions made up with 8% DMSO
Runx1 reverse	TCTCCGGGCCGAGTACCTTG			
Porcine Rag1 forward	CCTCTGTGTGCTGGTTCTGG	62°C	Amplifies TALEN target region in porcine Rag1	
Porcine Rag1 reverse	GCTTGAATGCTGGTTGAGGC			
Ovine Rag1 forward	CAAGATGCATCAGTGGGATG	61°C	Amplifies TALEN target region in ovine Rag1	
Ovine Rag1 reverse	GAGAGAATTCCCACAGATGC			
MSTN TALEN forward	GTCAAGGTAACAGACACACC	58°C	Amplifies TALEN target region in bovine and ovine MSTN	
MSTN TALEN reverse	CACCCACAGCGATCTACTAC			
Ovine MSTN CRISPR forward	CTCCTTGCGGTAGGAGAGTG		Amplifies CRISPR target region in ovine MSTN	Used with MSTN TALEN reverse primer
Bovine MSTN CRISPR forward	GTCCTTGAGGTAGGAGAGTG		Amplifies CRISPR target region in bovine MSTN	Used with MSTN TALEN reverse primer
ELD/KKR forward	TACAGGGGAGAGCACCTGG	N/A	Sequences FokI fragment to check ELD/KKR mutations	

Primer name	Primer sequence (5'-3')	Annealing temperature	Purpose	Notes
Ovine PPT1 forward	CCAAGACCCAAGAGGTGAGA	56°C	Amplifies CRISPR target region of ovine PPT1	
Ovine PPT1 reverse	GCATCACGGTTAGCATGAGA			
Ovine MSTN 3' UTR forward	GTCATGTGGGACATAAAAGC	60°C	Amplifies region of 3' UTR of ovine MSTN which contains the Texel G/A SNP	
Ovine MSTN 3' UTR reverse	CTAGAGTAGCAGCCTAACA			
MSTN CRISPR F1 gBlock T7 forward	TTAATACGACTCACTATAGGGCAGACACACCAAAAAGATCT	61°C	Amplifies sgRNA sequence from gBlocks and adds a T7 promoter at the 5'end	All forward primers used with universal gBlock reverse
MSTN CRISPR F2 gBlock T7 forward	TTAATACGACTCACTATAGGGCTGTCGTTACCCTCTAACTG			
MSTN CRISPR R1 gBlock T7 forward	TTAATACGACTCACTATAGGGACGACAGCATCGAGATTCTG			
MSTN CRISPR R2 gBlock T7 forward	TTAATACGACTCACTATAGGGCTTCAAAATCCACAGTTAGA			
CRISPR gBlock reverse	AAAAGCACCGACTCGGTGCC			
MSTN CRISPR F1 forward	CACCCAGACACACCAAAAAGATCT	N/A	Ligates MSTN CRISPR sgRNA F1 into pSpCas9(BB) vector	
MSTN CRISPR F1 reverse	AAACAGATCTTTTTGGTGTGTCTG			
MSTN CRISPR F2 forward	CACCCTGTCGTTACCCTCTAACTG		Ligates MSTN CRISPR sgRNA F2 into pSpCas9(BB) vector	
MSTN CRISPR F2 reverse	AAACCAGTTAGAGGGTAACGACAG			
MSTN CRISPR R1 forward	CACCACGACAGCATCGAGATTCTG		Ligates MSTN CRISPR sgRNA R1 into pSpCas9(BB) vector	
MSTN CRISPR R1 reverse	AAACCAGAATCTCGATGCTGTCGT			
MSTN CRISPR R2 forward	CACCCTTCAAAATCCACAGTTAGA		Ligates MSTN CRISPR sgRNA R2 into pSpCas9(BB) vector	
MSTN CRISPR R2 reverse	AAACTCTAACTGTGGATTTTGAAG			
Porcine eIF4A1 forward	TGAGGCATGACAGGAACTCC	62°C	Amplifies TALEN target region in porcine eIF4A1	
Porcine eIF4A1 reverse	AACACACACACACCAACAGG			
Ovine eIF4A1 forward	CATATCAATTCTGCAACAGA	58°C	Amplifies TALEN target region in ovine eIF4A1	
Ovine eIF4A1 reverse	ATGTAGAATTGACGGATACC			
Porcine eIF4G1 forward	GTCTGTAGCCCTAACTAGCC	62°C	Amplifies TALEN target region in porcine eIF4G1	
Porcine eIF4G1 reverse	GATCTTGCGTGGTTCCTTTC			

Primer name	Primer sequence (5'-3')	Annealing temperature	Purpose	Notes
Ovine eIF4G1 forward	AAGAGTGGTCAGTAGCTAGC	53°C	Amplifies TALEN target region in ovine eIF4G1	
Ovine eIF4G1 reverse	CACTGTCCCAATGATCTTGC			
Porcine IKBKG forward	CTCGTCCCTGCTGATCGGAC	62°C	Amplifies TALEN target region in porcine IKBKG	
Porcine IKBKG reverse	CTCAGTCTCGGAGCAGGAGA			
Ovine IKBKG forward	GCTGGCTGAGAAGAAGGAAT	62°C	Amplifies TALEN target region in ovine IKBKG	
Ovine IKBKG reverse	CATGTCAGGAGCCTGATACT			
Bovine Nanos2 forward	GGGTCTTTCCAGAGGTCACA	60°C	Amplifies CRISPR target region of bovine Nanos2	
Bovine Nanos2 reverse	GCAGACAAGAGTCCGGAGAG			

A.2. TALEN sequences

On the following pages are listed the target DNA sequences and full RVD sequences for all TALENs used in this work. DNA sequences directly bound by TALENs are shown 5' to 3' in blue, while spacer sequences are shown in red. Note that all right TALENs bind to complementary bases on the reverse DNA strand. Differences in binding sites between species are highlighted in bold.

Rag1 TALENs

		Left TALEN																	Spacer	Right TALEN																			
Rag1 A (sheep and pig)	DNA binding site	T	T	A	A	A	T	T	T	T	C	A	G	A	A	T	G	G	A	AGTTTAAGCTATTCAG	G	G	T	G	A	G	A	T	C	C	T	T	T	G	A	A	A	A
	RVD sequences	NG1	NG2	NI3	NI4	NI5	NG6	NG7	NG8	NG9	HD10	NI1	NN2	NI3	NI4	NG5	NN6	NN7	pLR_NI		pLR_HD	HD7	NI6	HD5	NG4	HD3	NG2	NI1	NN10	NN9	NI8	NI7	NI6	HD5	NG4	NG3	NG2	NG1
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7							pFUS_A									
	Destination vector	pRCIScript-GoldyTALEN																			pRCIScript-GoldyTALEN																	

		Left TALEN																	Spacer	Right TALEN																			
Rag1 B (pig)	DNA binding site	T	C	A	G	C	C	A	G	C	A	T	G	G	C	T	G	T	C	TCTTTGCCACCCACTCTG	G	G	A	C	T	C	A	G	T	T	C	**C**	G	C	C	C	C	A
	RVD sequences	NG1	HD2	NI3	NN4	HD5	HD6	NI7	NN8	HD9	NI10	NG1	NN2	NN3	HD4	NG5	NN6	NG7	pLR_HD		pLR_HD	HD7	NG6	NN5	NI4	NN3	NG2	HD1	NI10	NI9	NN8	**NN7**	HD6	NN5	NN4	NN3	NN2	NG1
Rag1 B (sheep)	DNA binding site	T	C	A	G	C	C	A	G	C	A	T	G	G	C	T	G	T	C	TCTTTGCCACCCACTCTG	G	G	A	C	T	C	A	G	T	T	C	**T**	G	C	C	C	C	A
	RVD sequences	Same as Rag1 pig B left TALEN																			pLR_HD	HD7	NG6	NN5	NI4	NN3	NG2	HD1	NI10	NI9	NN8	**NN7**	HD6	NN5	NN4	NN3	NN2	NG1
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7							pFUS_A									
	Destination vector	pRCIScript-GoldyTALEN																			pRCIScript-GoldyTALEN																	

		Left TALEN																	Spacer	Right TALEN																			
Rag1 C (pig)	DNA binding site	T	C	A	G	T	T	C	**C**	G	C	C	C	C	A	G	A	T	G	AAATCCAGCACCCCCACA	T	T	A	A	A	T	T	T	T	C	A	G	A	A	T	G	G	A
	RVD sequences	NG1	HD2	NI3	NN4	NG5	NG6	HD7	**HD8**	NN9	HD10	HD1	HD2	HD3	NI4	NN5	NI6	NG7	pLR_NN		pLR_NI	NI7	NG6	NG5	NG4	NI3	NI2	NI1	NI10	NN9	NG8	HD7	NG6	NG5	NI4	HD3	HD2	NG1
Rag1 C (sheep)	DNA binding site	T	C	A	G	T	T	C	**T**	G	C	C	C	C	A	G	A	T	G	AAATCCAGCACCCCCACA	T	T	A	A	A	T	T	T	T	C	A	G	A	A	T	G	G	A
	RVD sequences	NG1	HD2	NI3	NN4	NG5	NG6	HD7	**NG8**	NN9	HD10	HD1	HD2	HD3	NI4	NN5	NI6	NG7			Same as Rag1 pig C right TALEN																			
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7							pFUS_A									
	Destination vector	pRCIScript-GoldyTALEN																			pRCIScript-GoldyTALEN																	

Runx1 TALENs

		Left TALEN																		Spacer	Right TALEN																	
Runx1 A (pig)	DNA binding site	T	G	G	G	C	G	C	C	C	C	G	G	A	C	G	C	C	**A**	GCGCCGCCCTGGCCGG	C	A	A	G	C	T	G	A	G	G	A	G	C	G	G	C	G	A
	RVD sequences	NG1	NN2	NN3	NN4	HD5	NN6	HD7	HD8	HD9	HD10	NN1	NN2	NI3	HD4	NN5	HD6	HD7	**pLR_NI**		pLR_NN	NG7	NG6	HD5	NN4	NI3	HD2	NG1	HD10	HD9	NG8	HD7	NN6	HD5	HD4	NN3	HD2	NG1
Runx1 A (sheep)	DNA binding site	T	G	G	G	C	G	C	C	C	C	G	G	A	C	G	C	C	**G**	GCGCCGCCCTGGCCGG	C	A	A	G	C	T	G	A	G	G	A	G	C	G	G	C	G	A
	RVD sequences	NG1	NN2	NN3	NN4	HD5	NN6	HD7	HD8	HD9	HD10	NN1	NN2	NI3	HD4	NN5	HD6	HD7	**pLR_NN**		Same as Runx1 A pig right TALEN																	
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7								pFUS_A								
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR																

		Left TALEN																		Spacer	Right TALEN																	
Runx1 B (pig)	DNA binding site	T	G	A	G	G	A	G	C	G	G	C	G	A	C	C	G	C	A	GCATGGTGGAGGTGCT	G	G	C	C	G	A	C	C	A	C	C	C	**C**	G	G	C	G	A
	RVD sequences	NG1	NN2	NI3	NN4	NN5	NI6	NN7	HD8	NN9	NN10	HD1	NN2	NI3	HD4	HD5	NN6	HD7	pLR_NI		pLR_HD	HD7	NN6	NN5	HD4	NG3	NN2	NN1	NG10	NN9	NN8	NN7	**NN6**	HD5	HD4	NN3	HD2	NG1
Runx1 B (sheep)	DNA binding site	T	G	A	G	G	A	G	C	G	G	C	G	A	C	C	G	C	A	GCATGGTGGAGGTGCT	G	G	C	C	G	A	C	C	A	C	C	C	**G**	G	G	C	G	A
	RVD sequences	Same as Runx1 B pig left TALEN																				HD7	NN6	NN5	HD4	NG3	NN2	NN1	NG10	NN9	NN8	NN7	**HD6**	HD5	HD4	NN3	HD2	NG1
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7								pFUS_A								
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR																

		Left TALEN																		Spacer	Right TALEN																	
Runx1 C (pig)	DNA binding site	T	G	G	A	G	G	T	G	C	T	G	G	C	C	G	A	C	C	ACCCCGGCGAGCTGGT	G	C	G	C	A	C	C	G	A	C	A	G	**T**	C	C	C	A	A
	RVD sequences	NG1	NN2	NN3	NI4	NN5	NN6	NG7	NN8	HD9	NG10	NN1	NN2	HD3	HD4	NN5	NI6	HD7	pLR_HD		pLR_HD	NN7	HD6	NN5	NG4	NN3	NN2	HD1	NG10	NN9	NG8	HD7	**NI6**	NN5	NN4	NN3	NG2	NG1
Runx1 C (sheep)	DNA binding site	T	G	G	A	G	G	T	G	C	T	G	G	C	C	G	A	C	C	ACCCCGGCGAGCTGGT	G	C	G	C	A	C	C	G	A	C	A	G	**C**	C	C	C	A	A
	RVD sequences	Same as Runx1 C pig left TALEN																				NN7	HD6	NN5	NG4	NN3	NN2	HD1	NG10	NN9	NG8	HD7	**NN6**	NN5	NN4	NN3	NG2	NG1
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7								pFUS_A								
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR																

IKBKG TALENs

		Left TALEN																	Spacer	Right TALEN																	
IKBKG A (pig)	DNA binding site	T	C	A	T	C	A	C	T	C	C	T	T	T	**C**	A	C	C	**C**	GGCCCTGCCCAGCCAGAG	G	A	G	A	A	G	C	C	C	C	C	C	**C**	G	A	**G**	G	A
	RVD sequences	NG1	HD2	NI3	NG4	HD5	NI6	HD7	NG8	HD9	HD10	NG1	NG2	NG3	**HD4**	NI5	HD6	HD7	**pLR_HD**		pLR_HD	NG7	HD6	NG5	NG4	HD3	NN2	NN1	NN10	NN9	NN8	NN7	**NN6**	HD5	NG4	**HD3**	HD2	NG1
IKBKG A (sheep)	DNA binding site	T	C	A	T	C	A	C	T	C	C	T	T	T	**A**	A	C	C	**T**	GGCCCTGCCCAGCCAGAG	G	A	G	A	A	G	C	C	C	C	C	C	**T**	G	A	**C**	G	A
	RVD sequences	NG1	HD2	NI3	NG4	HD5	NI6	HD7	NG8	HD9	HD10	NG1	NG2	NG3	**NI4**	NI5	HD6	HD7	**pLR_NG**			NG7	HD6	NG5	NG4	HD3	NN2	NN1	NN10	NN9	NN8	NN7	**NI6**	HD5	NG4	**NN3**	HD2	NG1
	Subset vectors	pFUS_A										pFUS_B7									pFUS_B7							pFUS_A									
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																		pCAG-T7-TALEN(Sangamo)-FokI-KKR																	

		Left TALEN																	Spacer	Right TALEN																	
IKBKG B (pig)	DNA binding site	T	**G**	**C**	C	**C**	C	C	**C**	G	C	C	C	C	C	A	G	C	T	C	ATCACTCCTTTCACCCGG	C	C	C	T	G	C	C	C	A	**G**	C	C	A	G	A	G	G	A
	RVD sequences	NG1	**NN2**	**HD3**	HD4	**HD5**	HD6	HD7	**HD8**	NN9	HD10	HD1	HD2	HD3	NI4	NN5	HD6	NG7	pLR_HD		pLR_NN	NN7	NN6	NI5	HD4	NN3	NN2	NN1	NG10	**HD9**	NN8	NN7	NG6	HD5	NG4	HD3	HD2	NG1
IKBKG B (sheep)	DNA binding site	T	**A**	**A**	C	**A**	C	C	**T**	G	C	C	C	C	A	G	C	T	C	ATCACTCCTTTCACCCGG	C	C	C	T	G	C	C	C	A	**T**	C	C	A	G	A	G	G	A
	RVD sequences	NG1	**NI2**	**NI3**	HD4	**NI5**	HD6	HD7	**NG8**	NN9	HD10	HD1	HD2	HD3	NI4	NN5	HD6	NG7			pLR_NN	NN7	NN6	NI5	HD4	NN3	NN2	NN1	NG10	**NI9**	NN8	NN7	NG6	HD5	NG4	HD3	HD2	NG1
	Subset vectors	pFUS_A										pFUS_B7									pFUS_B7							pFUS_A									
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																		pCAG-T7-TALEN(Sangamo)-FokI-KKR																	

		Left TALEN																	Spacer	Right TALEN																	
IKBKG C (pig)	DNA binding site	T	T	T	**C**	A	C	C	**C**	G	G	C	C	C	T	G	C	C	C	AGCCAGAGGAGAAGCC	C	C	C	C	**C**	G	A	**G**	G	A	G	C	**C**	C	C	C	A
	RVD sequences	NG1	NG2	NG3	**HD4**	NI5	HD6	HD7	**HD8**	NN9	NN10	HD1	HD2	HD3	NG4	NN5	HD6	HD7	pLR_HD		pLR_NN	NN7	NN6	NN5	**NN4**	HD3	NG2	**HD1**	HD10	NG9	HD8	NN7	NN6	**NN5**	NN4	NN3	NN2	NG1
IKBKG C (sheep)	DNA binding site	T	T	T	**A**	A	C	C	**T**	G	G	C	C	C	T	G	C	C	C	AGCCAGAGGAGAAGCC	C	C	C	C	**T**	G	A	**C**	G	A	G	C	**G**	C	C	C	A
	RVD sequences	NG1	NG2	NG3	**NI4**	NI5	HD6	HD7	**NG8**	NN9	NN10	HD1	HD2	HD3	NG4	NN5	HD6	HD7				NN7	NN6	NN5	**NI4**	HD3	NG2	**NN1**	HD10	NG9	HD8	NN7	NN6	**HD5**	NN4	NN3	NN2	NG1
	Subset vectors	pFUS_A										pFUS_B7									pFUS_B7							pFUS_A									
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																		pCAG-T7-TALEN(Sangamo)-FokI-KKR																	

eIF4G1 TALENs

		Left TALEN																	Spacer	Right TALEN																	
eIF4G1 A (pig)	DNA binding sites	T	C	C	**G**	T	C	C	T	T	T	G	C	C	A	A	C	C	T	TGGCCGACCAGCCCTT	A	G	C	A	A	C	C	G	T	G	G	G	C	C	C	C	C	**A**
	RVD sequences	NG1	HD2	HD3	**NN4**	NG5	HD6	HD7	NG8	NG9	NG10	NN1	HD2	HD3	NI4	NI5	HD6	HD7	pLR_NG		pLR_NG	HD7	NN6	NG5	NG4	NN3	NN2	HD1	NI10	HD9	HD8	HD7	NN6	NN5	NN4	NN3	NN2	**NG1**
eIF4G1 A (sheep)	DNA binding site	T	C	C	**A**	T	C	C	T	T	T	G	C	C	A	A	C	C	T	TGGCCGACCAGCCCTT	A	G	C	A	A	C	C	G	T	G	G	G	C	C	C	C	C	**G**
	RVD sequences	NG1	HD2	HD3	**NI4**	NG5	HD6	HD7	NG8	NG9	NG10	NN1	HD2	HD3	NI4	NI5	HD6	HD7				HD7	NN6	NG5	NG4	NN3	NN2	HD1	NI10	HD9	HD8	HD7	NN6	NN5	NN4	NN3	NN2	**HD1**
	Subset vectors	pFUS_A										pFUS_B7									pFUS_B7							pFUS_A									
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR															

		Left TALEN																	Spacer	Right TALEN																	
eIF4G1 B (sheep and pig)	DNA binding site	T	G	G	C	C	G	A	C	C	A	G	C	C	C	T	T	A	G	CAACCGTGGGCCCCCAAG	G	G	G	T	G	G	G	C	C	A	G	G	T	G	G	G	G	A
	RVD sequences	NG1	NN2	NN3	HD4	HD5	NN6	NI7	HD8	HD9	NI10	NN1	HD2	HD3	HD4	NG5	NG6	NI7	pLR_NN		pLR_HD	HD7	HD6	NI5	HD4	HD3	HD2	NN1	NN10	NG9	HD8	HD7	NI6	HD5	HD4	HD3	HD2	NG1
	Subset vectors	pFUS_A										pFUS_B7									pFUS_B7							pFUS_A									
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR															

		Left TALEN																	Spacer	Right TALEN																	
eIF4G1 C (sheep and pig)	DNA binding site	T	G	T	G	G	C	C	C	A	G	A	C	T	T	C	A	C	T	CCGTCCTTTGCCAACC	T	T	G	G	C	C	G	A	C	C	A	G	C	C	C	T	T	A
	RVD sequences	NG1	NN2	NG3	NN4	NN5	HD6	HD7	HD8	NI9	NN10	NI1	HD2	NG3	NG4	HD5	NI6	HD7	pLR_NG		pLR_NI	NI7	HD6	HD5	NN4	NN3	HD2	NG1	NN10	NN9	NG8	HD7	NN6	NN5	NN4	NI3	NI2	NG1
	Subset vectors	pFUS_A										pFUS_B7									pFUS_B7							pFUS_A									
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR															

eIF4A1 TALENs

		Left TALEN																	Spacer	Right TALEN																	
eIF4A1 A (pig)	DNA binding site	**T**	G	G	T	A	C	C	A	A	T	**G**	T	G	C	G	T	G	C	TGAGGTGCAGAAGCTAC	A	G	A	T	G	G	A	A	G	C	**T**	C	C	C	C	A	**C**	A
	RVD sequences	**NG1**	NN2	NN3	NG4	NI5	HD6	HD7	NI8	NI9	NG10	**NN1**	NG2	NN3	HD4	NN5	NG6	NN7	pLR-HD		pLR-NG	HD7	NG6	NI5	HD4	HD3	NG2	NG1	HD10	NN9	**NI8**	NN7	NN6	NN5	NN4	NG3	**NN2**	NG1
eIF4A1 A (sheep)	DNA binding site	**G**	G	G	T	A	C	C	A	A	T	**A**	T	G	C	G	T	G	C	TGAGGTGCAGAAGCTAC	A	G	A	T	G	G	A	A	G	C	**C**	C	C	C	C	A	**T**	A
	RVD sequences	**NN1**	NN2	NN3	NG4	NI5	HD6	HD7	NI8	NI9	NG10	**NI1**	NG2	NN3	HD4	NN5	NG6	NN7				HD7	NG6	NI5	HD4	HD3	NG2	NG1	HD10	NN9	**NN8**	NN7	NN6	NN5	NN4	NG3	**NI2**	NG1
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7							pFUS_A								
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR															

		Left TALEN																	Spacer	Right TALEN																	
eIF4A1 B (pig)	DNA binding site	T	G	C	C	T	G	C	A	T	T	G	G	**T**	G	G	T	A	C	CAATGTGCGTGCTGAGGT	G	C	A	G	A	A	G	C	T	**A**	C	A	G	A	T	G	G	A
	RVD sequences	NG1	NN2	HD3	HD4	NG5	NN6	HD7	NI8	NG9	NG10	NN1	NN2	**NG3**	NN4	NN5	NG6	NI7	pLR-HD		pLR-NN	NN7	NG6	HD5	NG4	NG3	HD2	NN1	NI10	**NG9**	NN8	NG7	HD6	NG5	NI4	HD3	HD2	NG1
eIF4A1 B (sheep)	DNA binding site	T	G	C	C	T	G	C	A	T	T	G	G	**G**	G	G	T	A	C	CAATGTGCGTGCTGAGGT	G	C	A	G	A	A	G	C	T	**G**	C	A	G	A	T	G	G	A
	RVD sequences	NG1	NN2	HD3	HD4	NG5	NN6	HD7	NI8	NG9	NG10	NN1	NN2	**NN3**	NN4	NN5	NG6	NI7				NN7	NG6	HD5	NG4	NG3	HD2	NN1	NI10	**NN9**	NN8	NG7	HD6	NG5	NI4	HD3	HD2	NG1
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7							pFUS_A								
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR															

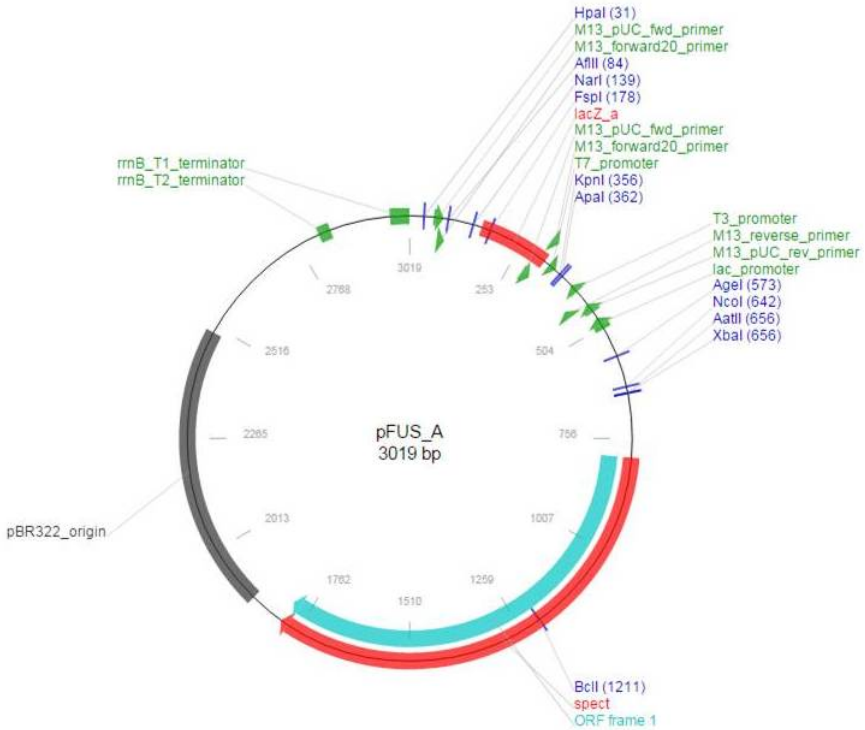
		Left TALEN																	Spacer	Right TALEN																	
eIF4A1 C (pig)	DNA binding site	T	G	C	**T**	T	C	**C**	T	G	C	C	A	T	G	C	C	T	G	CATTGGTGGTACCAATG	T	**G**	C	G	T	G	C	T	G	A	G	G	T	G	C	A	G	A
	RVD sequences	NG1	NN2	HD3	**NG4**	NG5	HD6	**HD7**	NG8	NN9	HD10	HD1	NI2	NG3	NN4	HD5	HD6	NG7	pLR-NN		pLR-NI	**HD7**	NN6	HD5	NI4	HD3	NN2	NI1	HD10	NG9	HD8	HD7	NI6	HD5	NN4	NG3	HD2	NG1
eIF4A1 C (sheep)	DNA binding site	T	G	C	**C**	T	C	**G**	T	G	C	C	A	T	G	C	C	T	G	CATTGGTGGTACCAATG	T	**A**	C	G	T	G	C	T	G	A	G	G	T	G	C	A	G	A
	RVD sequences	NG1	NN2	HD3	**HD4**	NG5	HD6	**NN7**	NG8	NN9	HD10	HD1	NI2	NG3	NN4	HD5	HD6	NG7				**NG7**	NN6	HD5	NI4	HD3	NN2	NI1	HD10	NG9	HD8	HD7	NI6	HD5	NN4	NG3	HD2	NG1
	Subset vectors	pFUS_A										pFUS_B7										pFUS_B7							pFUS_A								
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																			pCAG-T7-TALEN(Sangamo)-FokI-KKR															

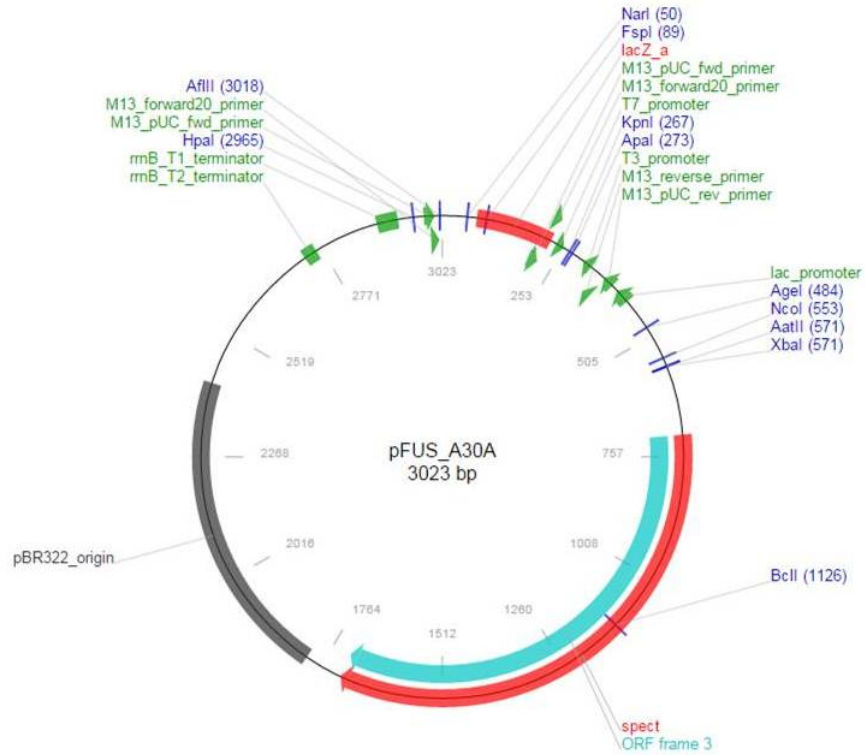
Myostatin TALENs

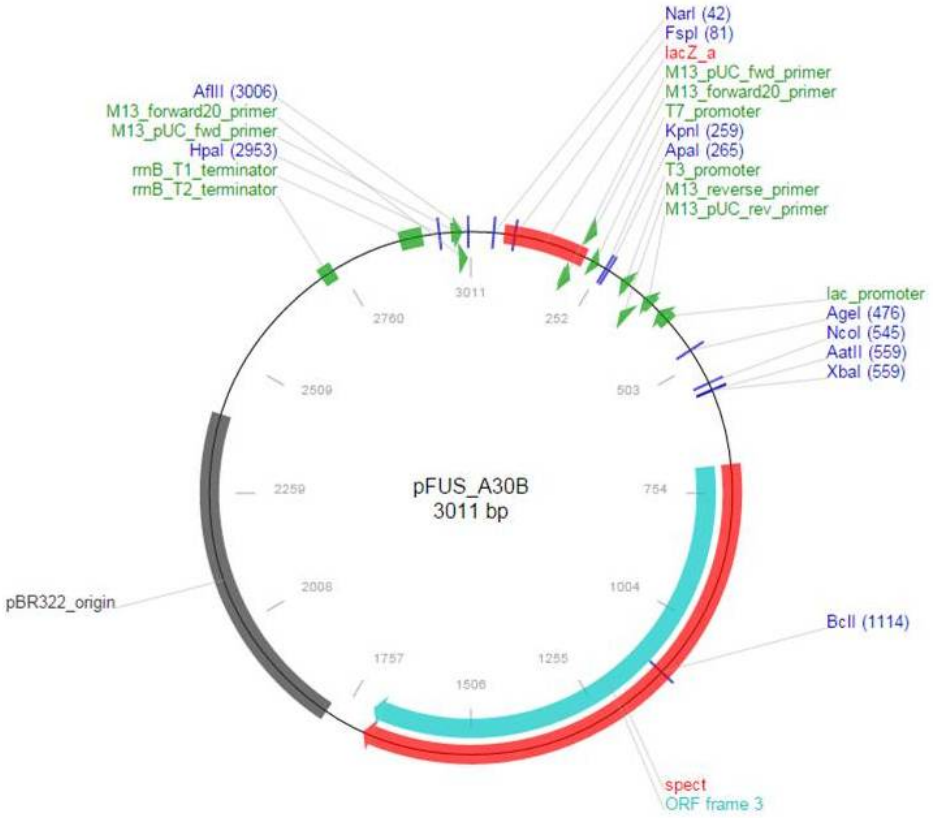
		Left TALEN																						Spacer	Right TALEN																			
MSTN (sheep)	DNA binding site	G	T	G	A	T	G	A	**G**	C	A	C	T	C	C	A	C	A	G	A	A	T	C	T	CGATGCTGTCGTTACC	C	T	C	T	A	A	C	T	G	T	G	G	A	T	T	T	T	G	A
	RVD sequences	NN1	NG2	NN3	NI4	NG5	NN6	NI7	**NN8**	HD9	NI10	HD1	NG2	HD3	HD4	NI5	HD6	NI7	NN8	NI9	NI10	NG1	HD2	pLR_NG		pLR_NN	NI8	NN7	NI6	NG5	NG4	NN3	NI2	HD1	NI10	HD9	HD8	NG7	NI6	NI5	NI4	NI3	HD2	NG1
MSTN (cow)	DNA binding site	G	T	G	A	T	G	A	**A**	C	A	C	T	C	C	A	C	A	G	A	A	T	C	T	CGATGCTGTCGTTACC	C	T	C	T	A	A	C	T	G	T	G	G	A	T	T	T	T	G	A
	RVD sequences	NN1	NG2	NN3	NI4	NG5	NN6	NI7	**NI8**	HD9	NI10	HD1	NG2	HD3	HD4	NI5	HD6	NI7	NN8	NI9	NI10	NG1	HD2	pLR_NG		Same as GDF8 sheep right TALEN																			
	Subset vectors	pFUSA30A										pFUSA30B										pFUS_B2				pFUS_B8									pFUS_A									
	Destination vector	pCAG-T7-TALEN(Sangamo)-FokI-ELD																							pCAG-T7-TALEN(Sangamo)-FokI-KKR																			

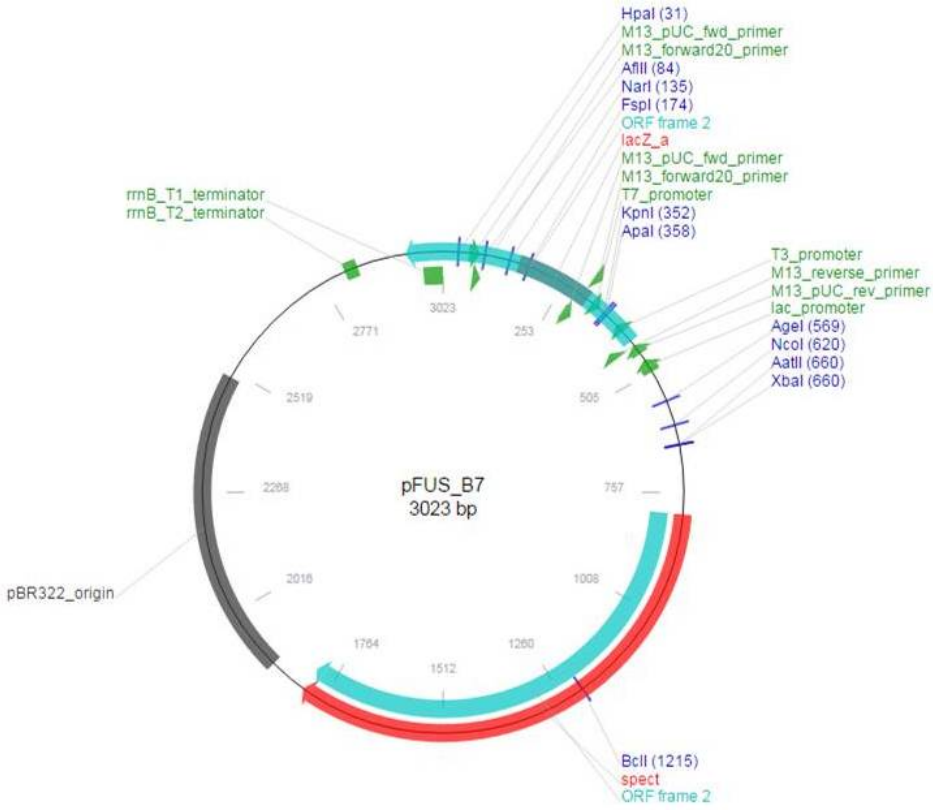
A.3. Plasmid maps

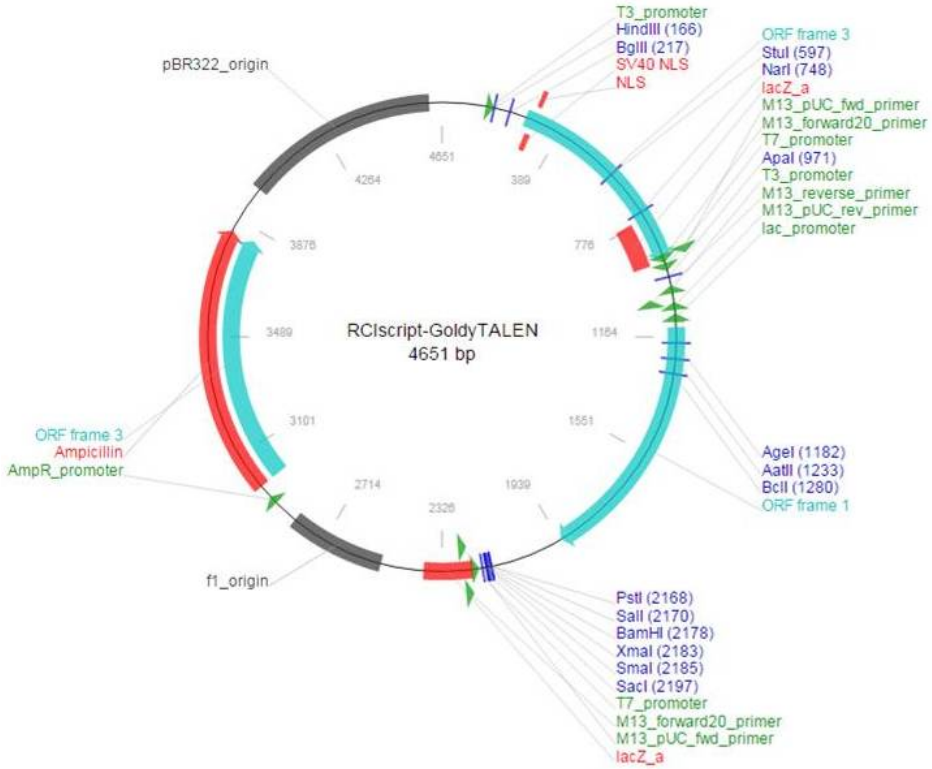
Plasmid maps for all TALEN and CRISPR/Cas9 vectors used in this work are shown below. Maps are taken from the Addgene website (`www.addgene.org`)

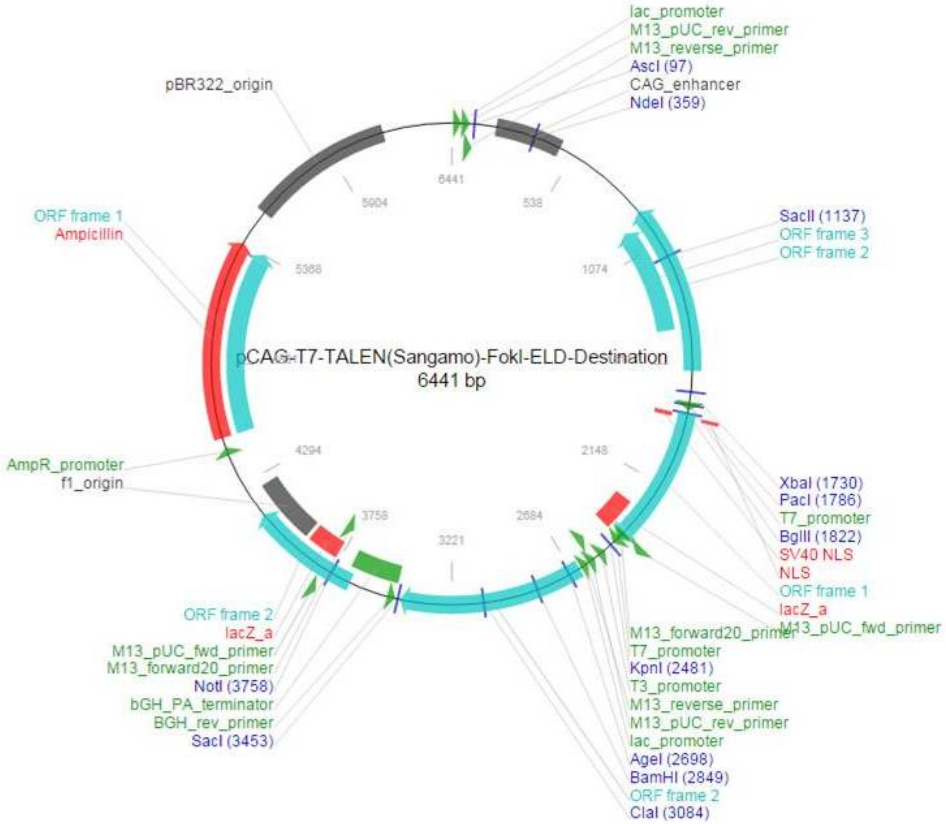


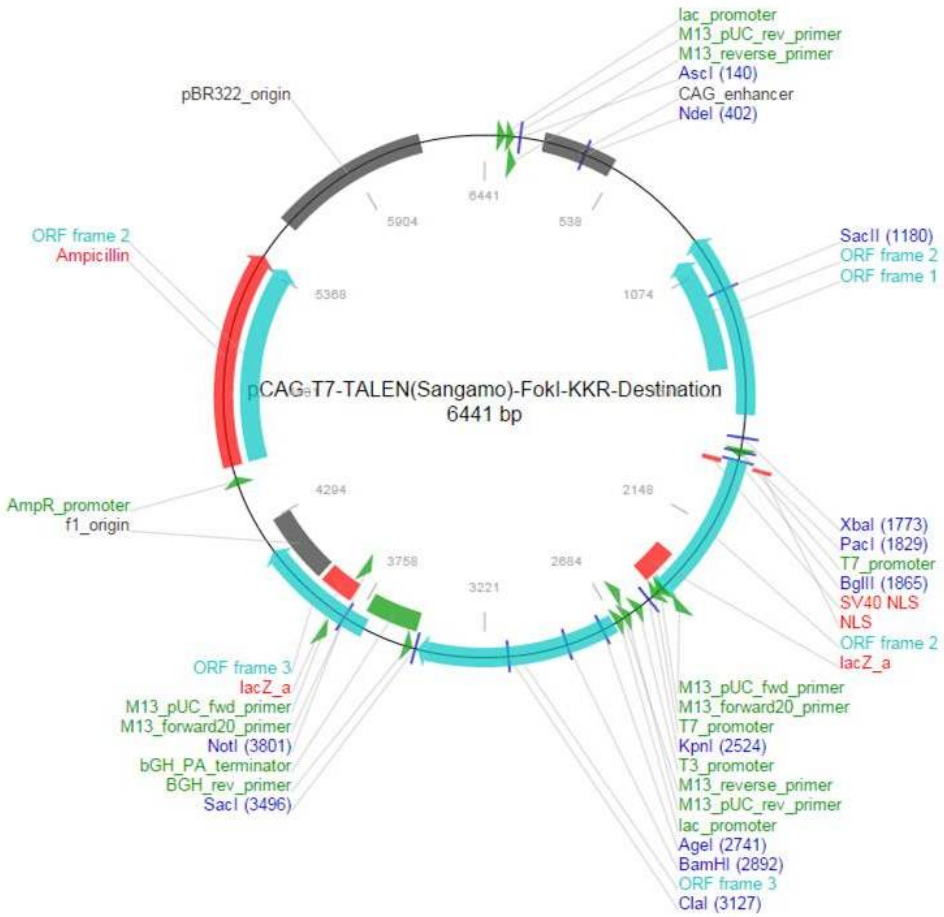


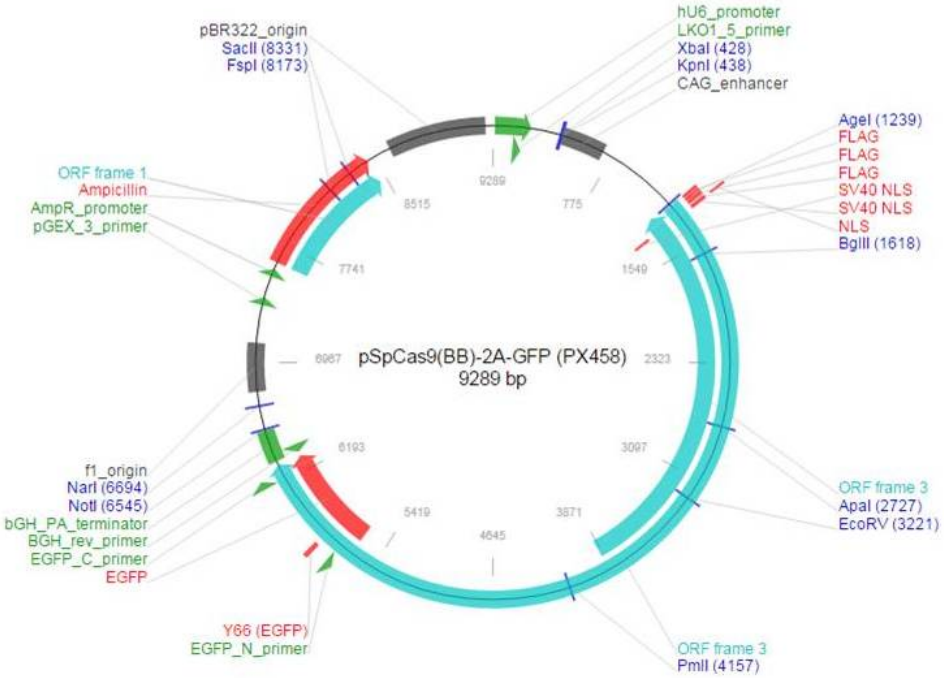

NarI (42)
FspI (81)
lacZ_a
M13_pUC_fwd_primer
M13_forward20_primer
T7_promoter
KpnI (259)
ApaI (265)
T3_promoter
M13_reverse_primer
M13_pUC_rev_primer
AflII (3006)
M13_forward20_primer
M13_pUC_fwd_primer
HpaI (2953)
rrnB_T1_terminator
rrnB_T2_terminator
lac_promoter
AgeI (476)
NcoI (545)
AatII (559)
XbaI (559)
pFUS_A30B
3011 bp
pBR322_origin
BclI (1114)
spect
ORF frame 3


HpaI (31)
M13_pUC_fwd_primer
M13_forward20_primer
AflII (84)
NarI (135)
FspI (174)
ORF frame 2
lacZ_a
M13_pUC_fwd_primer
M13_forward20_primer
T7_promoter
KpnI (352)
ApaI (358)
rrnB_T1_terminator
rrnB_T2_terminator
T3_promoter
M13_reverse_primer
M13_pUC_rev_primer
lac_promoter
AgeI (569)
NcoI (620)
AatII (660)
XbaI (660)
pFUS_B7
3023 bp
pBR322_origin
BclI (1215)
spect
ORF frame 2


RCIscript-GoldyTALEN
4651 bp
T3_promoter
HindIII (166)
BglII (217)
SV40 NLS
NLS
ORF frame 3
StuI (597)
NarI (748)
lacZ_a
M13_pUC_fwd_primer
M13_forward20_primer
T7_promoter
ApaI (971)
T3_promoter
M13_reverse_primer
M13_pUC_rev_primer
lac_promoter
AgeI (1182)
AatII (1233)
BclI (1280)
ORF frame 1
PstI (2168)
SalI (2170)
BamHI (2178)
XmaI (2183)
SmaI (2185)
SacI (2197)
T7_promoter
M13_forward20_primer
M13_pUC_fwd_primer
lacZ_a
f1_origin
ORF frame 3
Ampicillin
AmpR_promoter
pBR322_origin
4651
389
776
1164
1551
1939
2326
2714
3101
3489
3876
4264


pCAG-T7-TALEN(Sangamo)-FokI-ELD-Destination
6441 bp
lac_promoter
M13_pUC_rev_primer
M13_reverse_primer
AscI (97)
CAG_enhancer
NdeI (359)
SacII (1137)
ORF frame 3
ORF frame 2
XbaI (1730)
PacI (1786)
T7_promoter
BglII (1822)
SV40 NLS
NLS
ORF frame 1
lacZ_a
M13_pUC_fwd_primer
M13_forward20_primer
T7_promoter
KpnI (2481)
T3_promoter
M13_reverse_primer
M13_pUC_rev_primer
lac_promoter
AgeI (2698)
BamHI (2849)
ORF frame 2
ClaI (3084)
ORF frame 2
lacZ_a
M13_pUC_fwd_primer
M13_forward20_primer
NotI (3758)
bGH_PA_terminator
BGH_rev_primer
SacI (3453)
AmpR_promoter
f1_origin
ORF frame 1
Ampicillin
pBR322_origin
6441
538
1074
1611
2148
2684
3221
3758
4294
4831
5368
5904


lac_promoter
M13_pUC_rev_primer
M13_reverse_primer
AscI (140)
CAG_enhancer
NdeI (402)
pBR322_origin
6441
5904
538
SacII (1180)
ORF frame 2
ORF frame 1
ORF frame 2
Ampicillin
5368
1074
pCAG-T7-TALEN(Sangamo)-FokI-KKR-Destination
6441 bp
4294
2148
AmpR_promoter
f1_origin
XbaI (1773)
PacI (1829)
T7_promoter
BglII (1865)
SV40 NLS
NLS
ORF frame 2
lacZ_a
3758
2684
3221
ORF frame 3
lacZ_a
M13_pUC_fwd_primer
M13_forward20_primer
NotI (3801)
bGH_PA_terminator
BGH_rev_primer
SacI (3496)
M13_pUC_fwd_primer
M13_forward20_primer
T7_promoter
KpnI (2524)
T3_promoter
M13_reverse_primer
M13_pUC_rev_primer
lac_promoter
AgeI (2741)
BamHI (2892)
ORF frame 3
ClaI (3127)


hU6_promoter
LKO1_5_primer
XbaI (428)
KpnI (438)
CAG_enhancer
pBR322_origin
SacII (8331)
FspI (8173)
AgeI (1239)
FLAG
FLAG
FLAG
SV40 NLS
SV40 NLS
NLS
BglII (1618)
9289
8515
775
ORF frame 1
Ampicillin
AmpR_promoter
pGEX_3_primer
7741
1549
6967
pSpCas9(BB)-2A-GFP (PX458)
9289 bp
2323
6193
3097
f1_origin
NarI (6694)
NotI (6545)
bGH_PA_terminator
BGH_rev_primer
EGFP_C_primer
EGFP
5419
3871
4645
ORF frame 3
ApaI (2727)
EcoRV (3221)
Y66 (EGFP)
EGFP_N_primer
ORF frame 3
PmlI (4157)

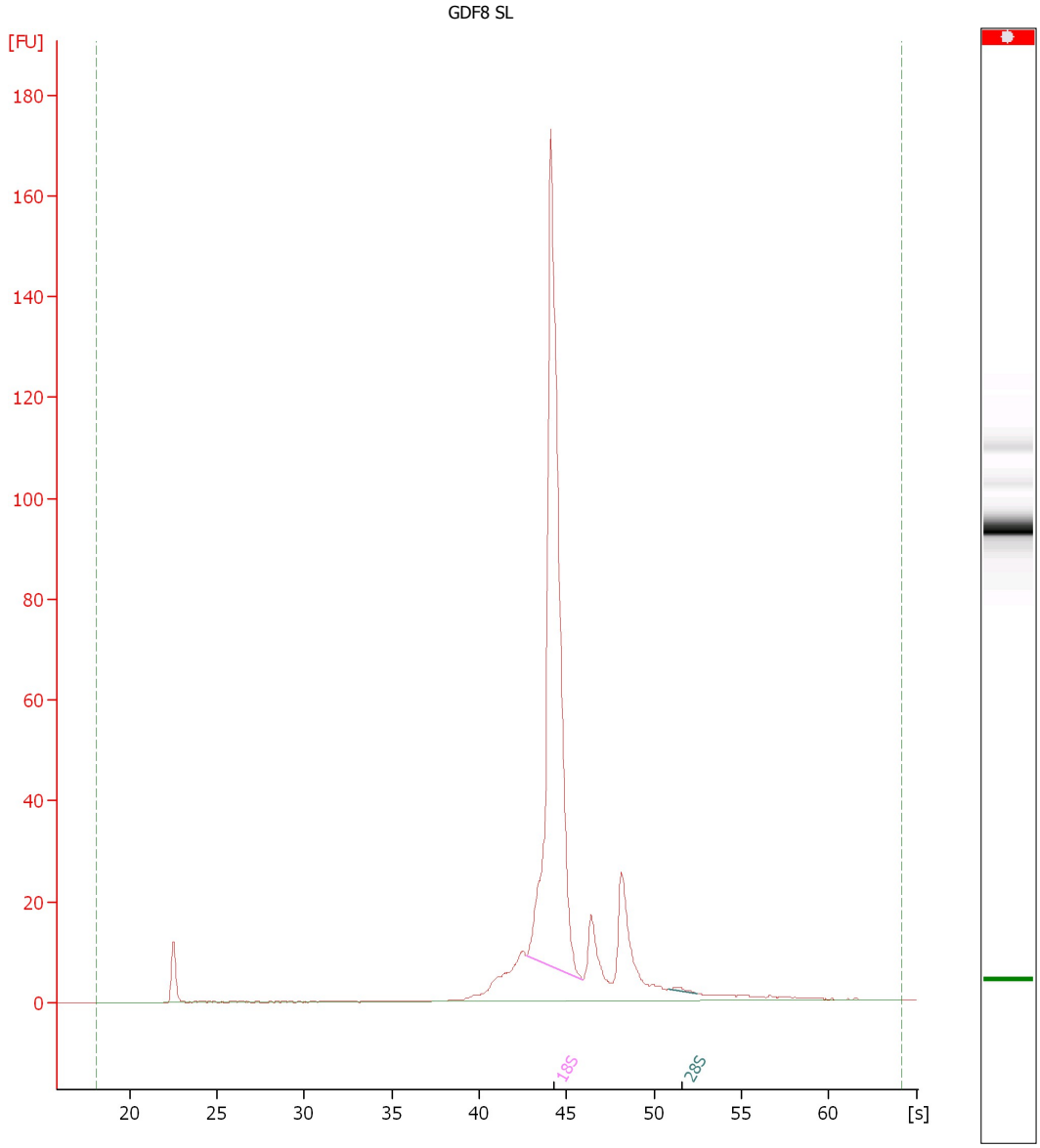
A.4. Agilent analysis of TALEN RNAs and sgRNAs for microinjection

MSTN TALENs

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

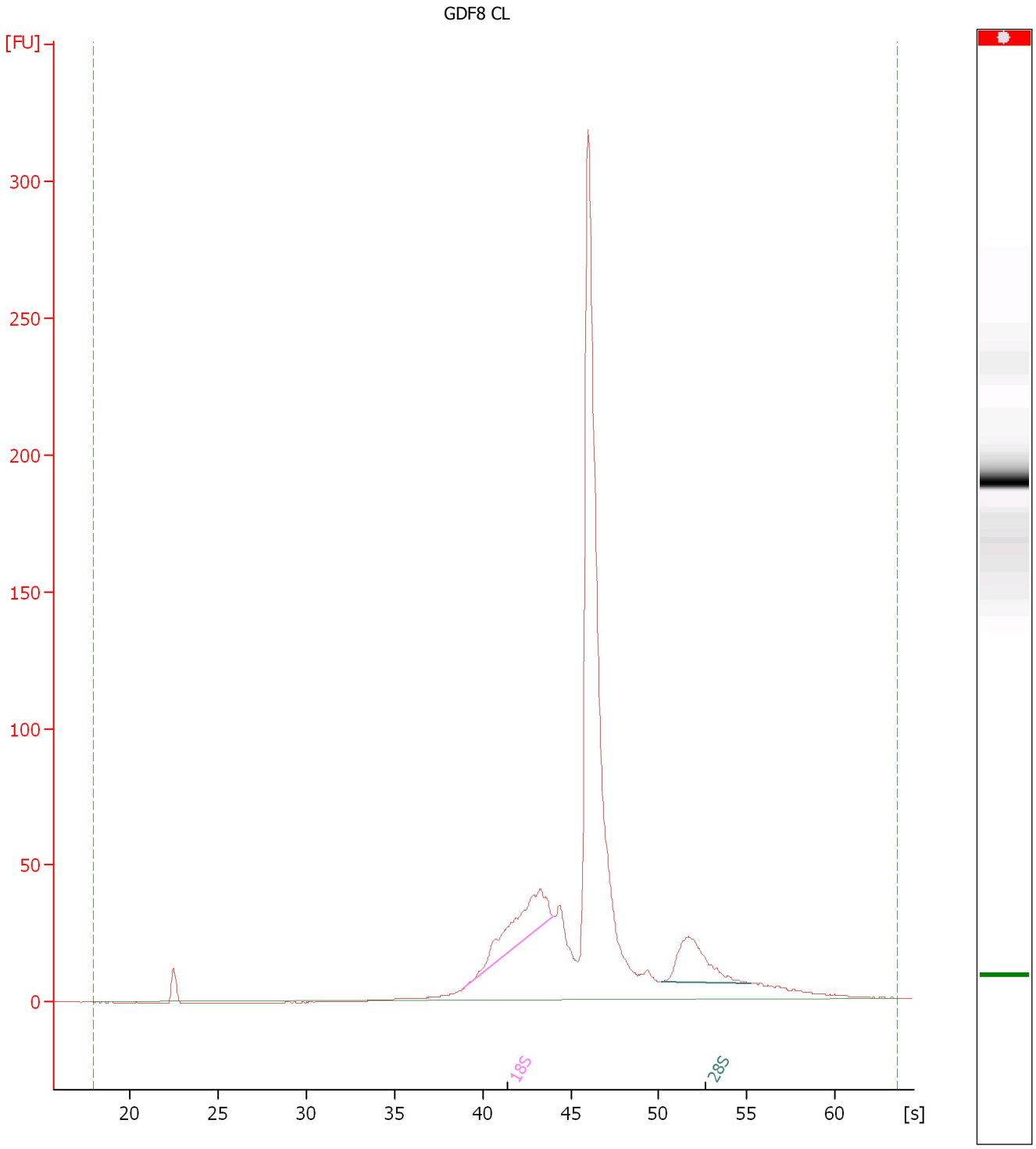
Electropherogram Summary



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

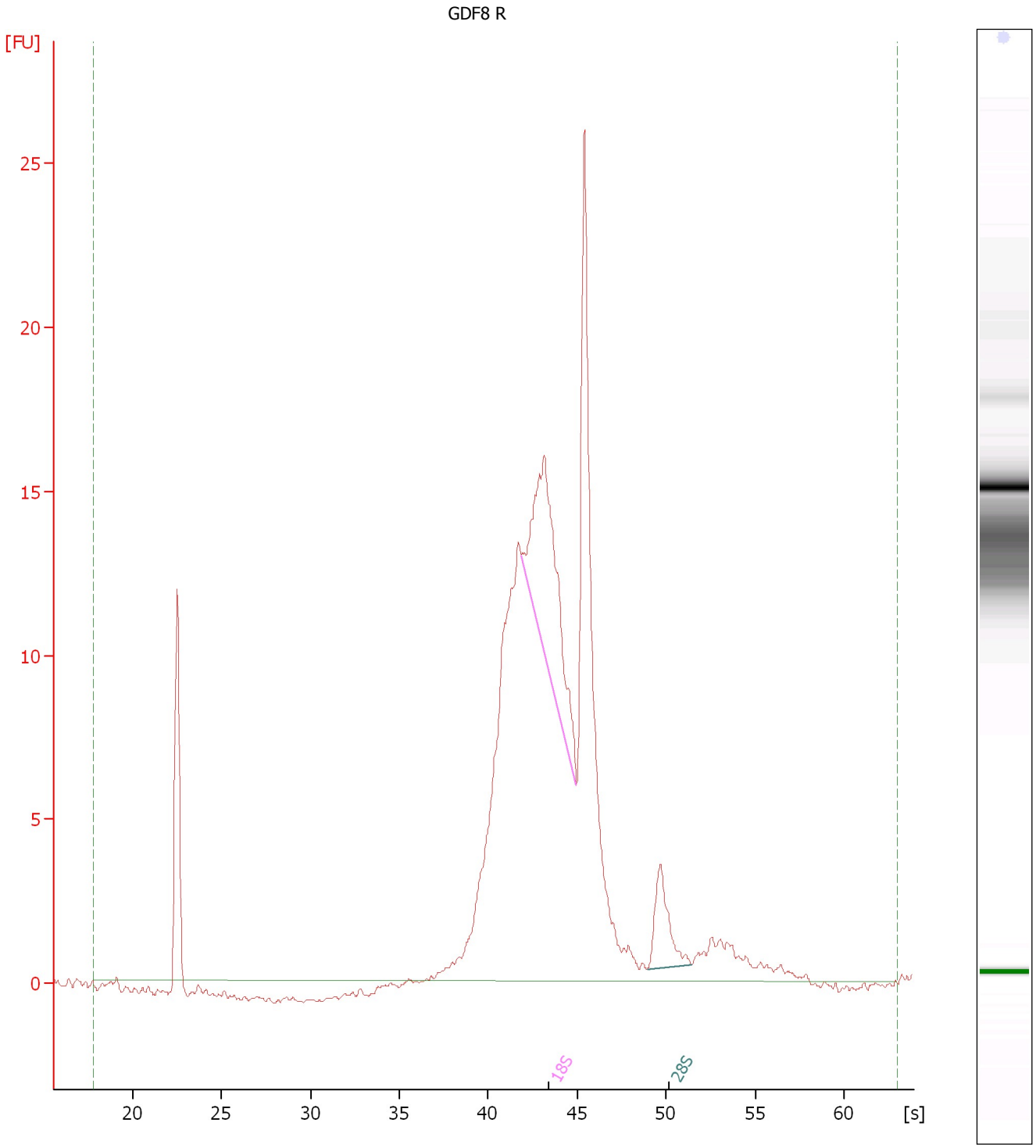
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

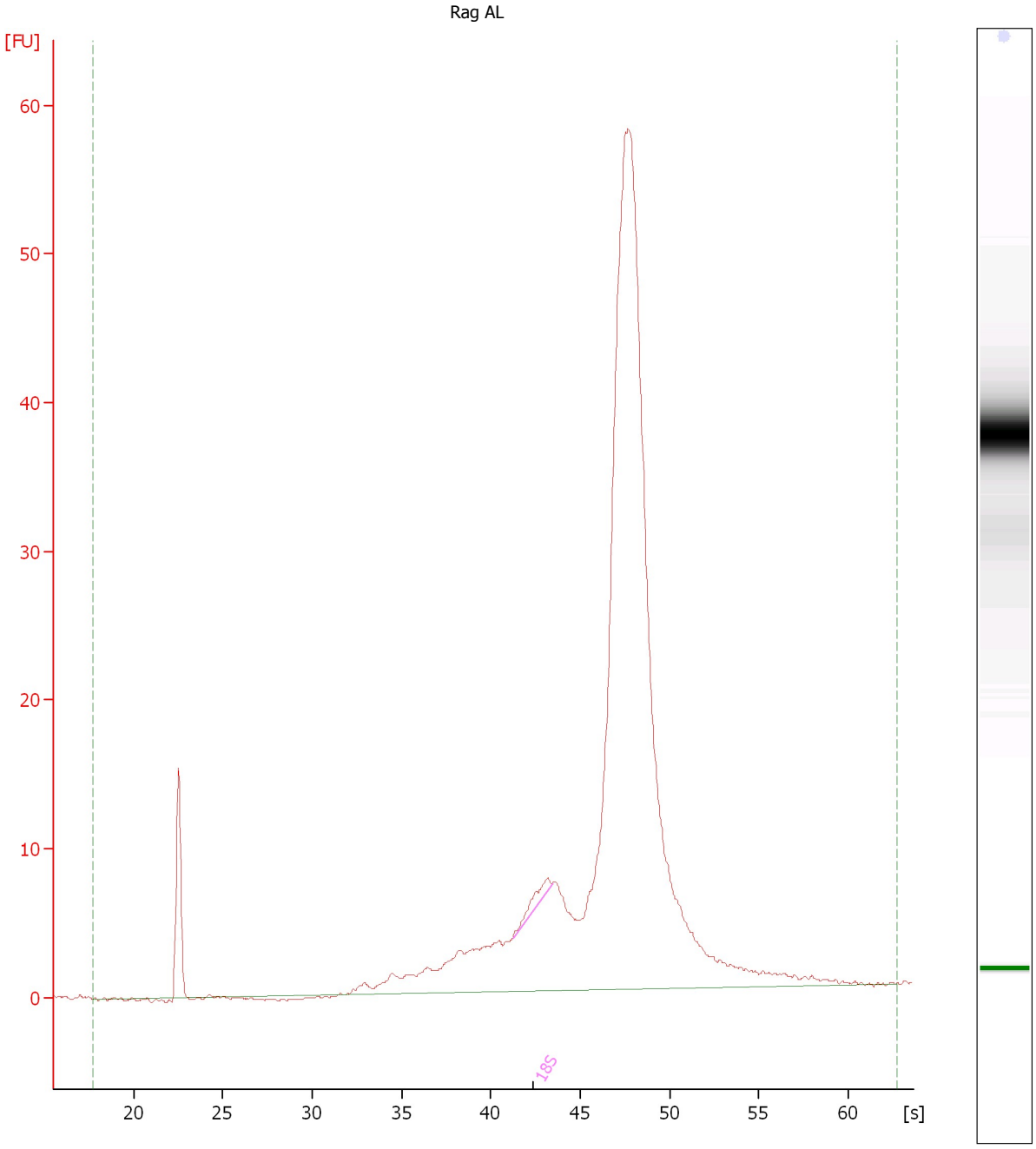
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

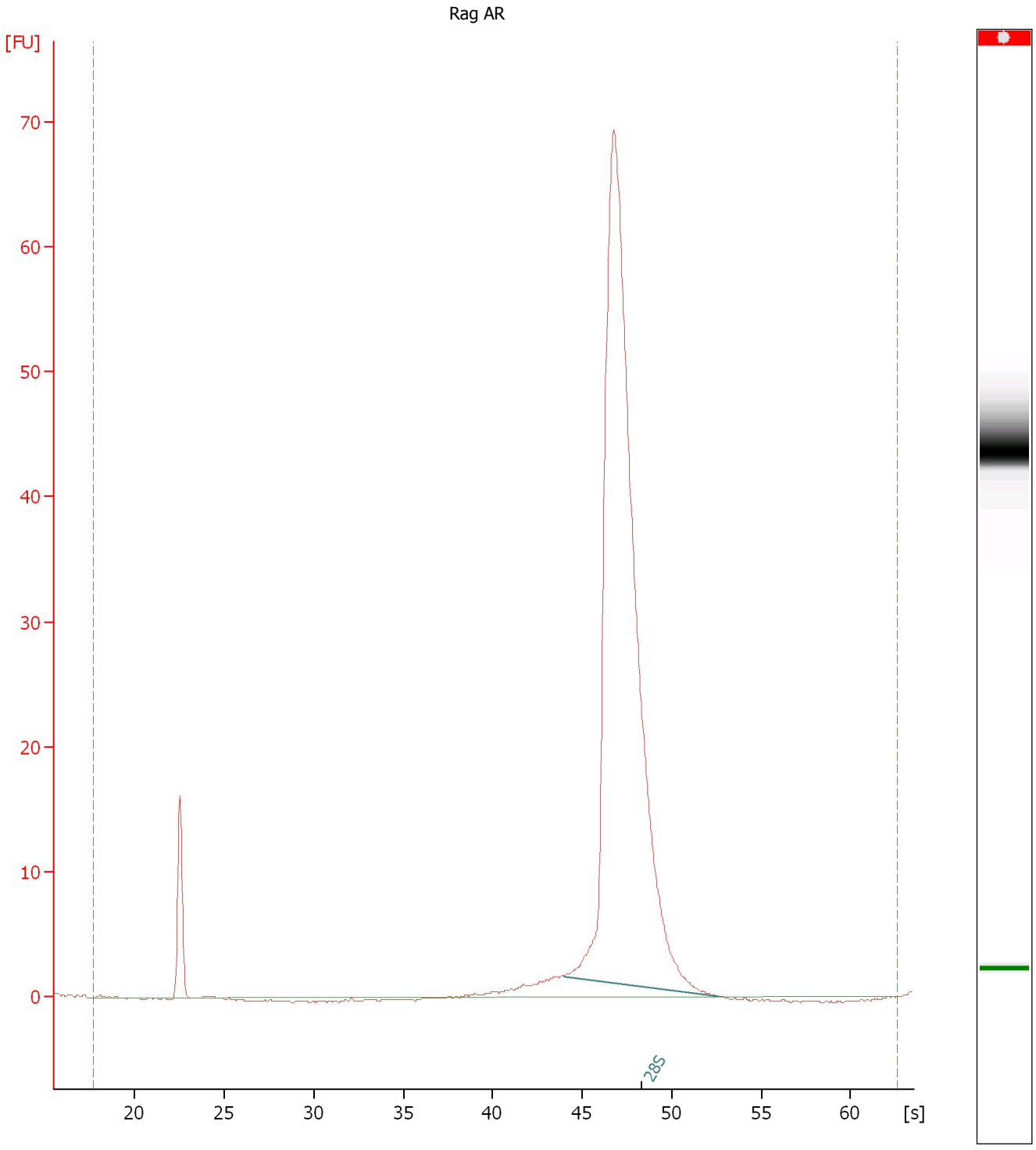
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

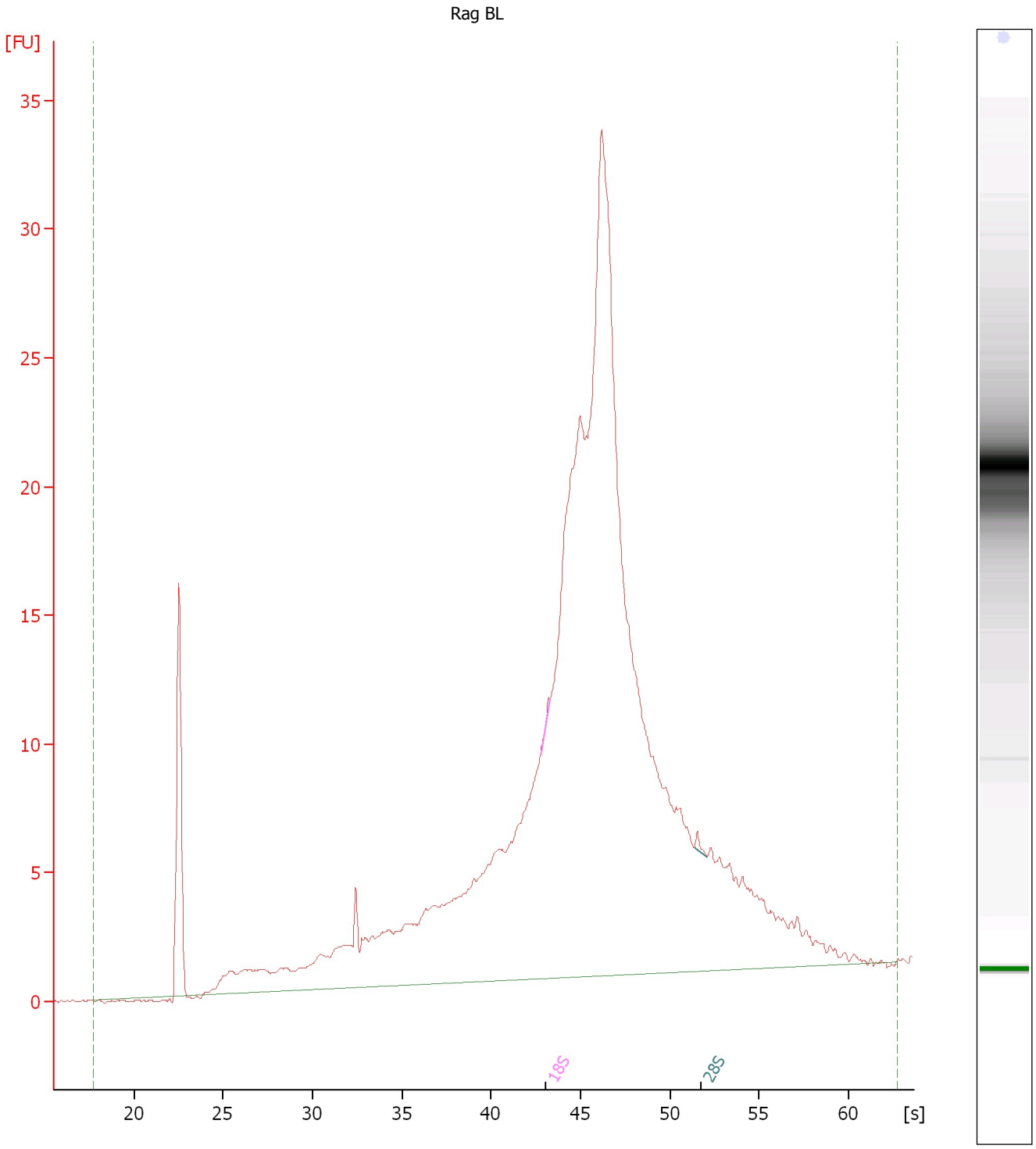
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

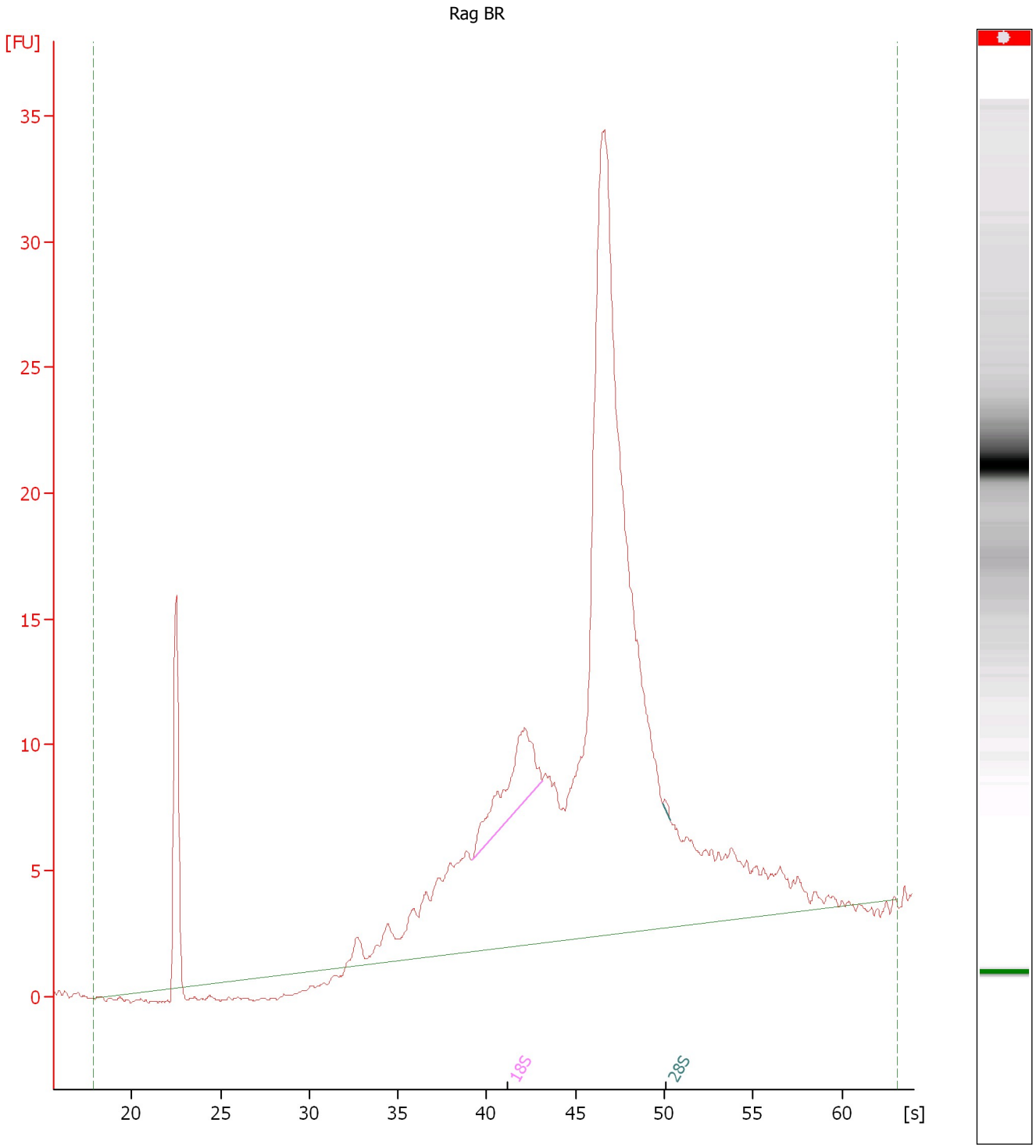
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

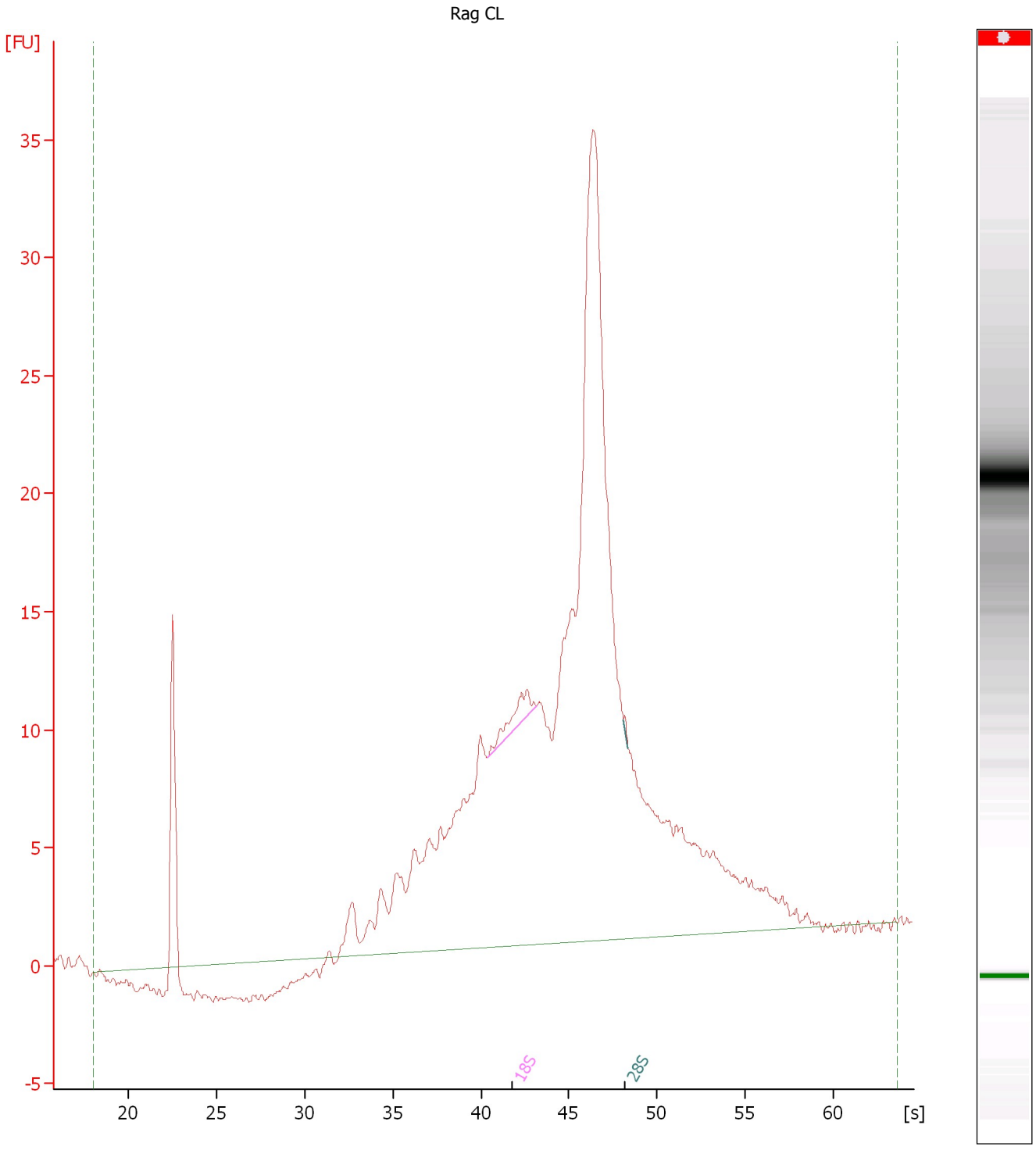
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

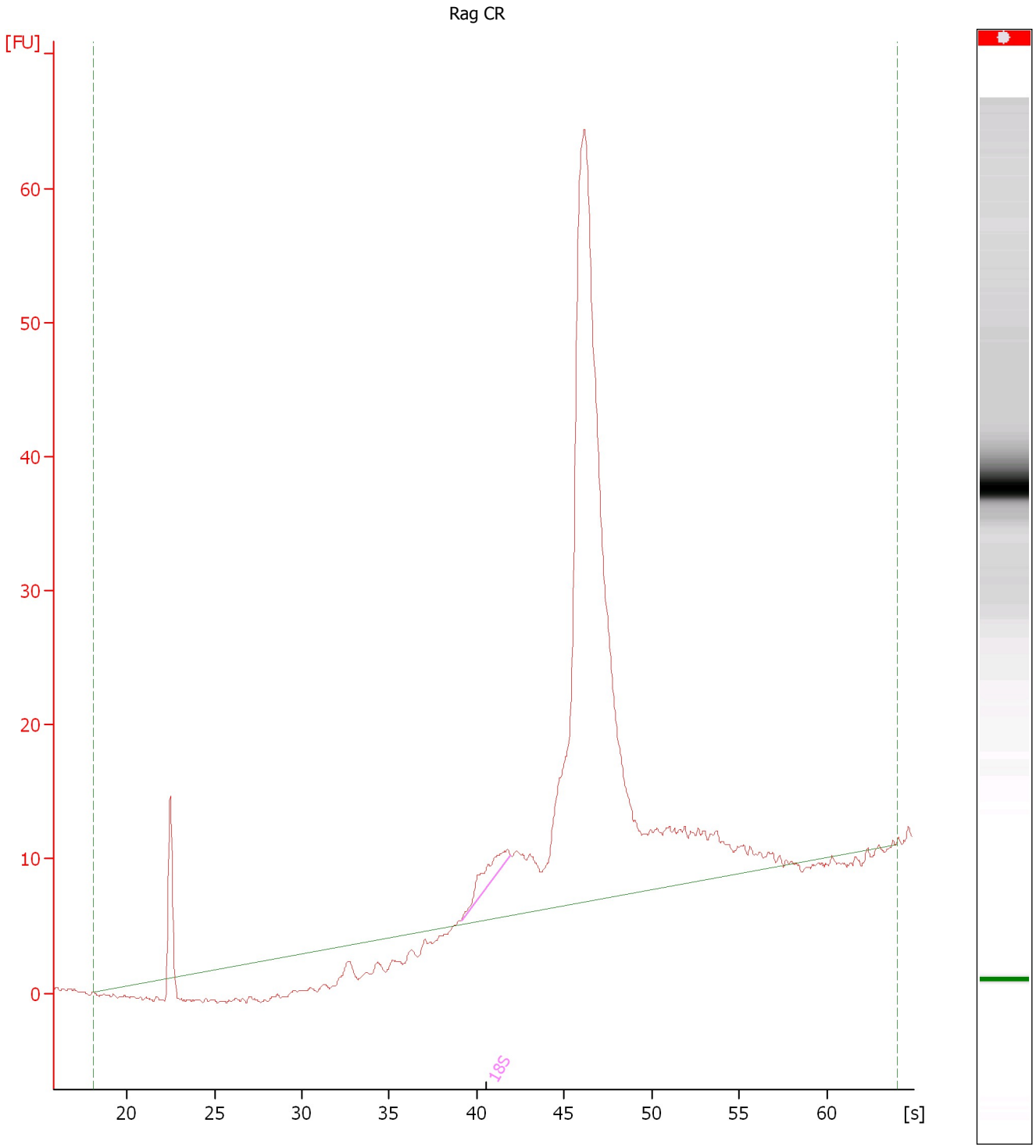
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-19_13-08-29.xad

Created: 19/01/2014 13:08:29
Modified: 19/01/2014 13:28:16

Electropherogram Summary Continued ...

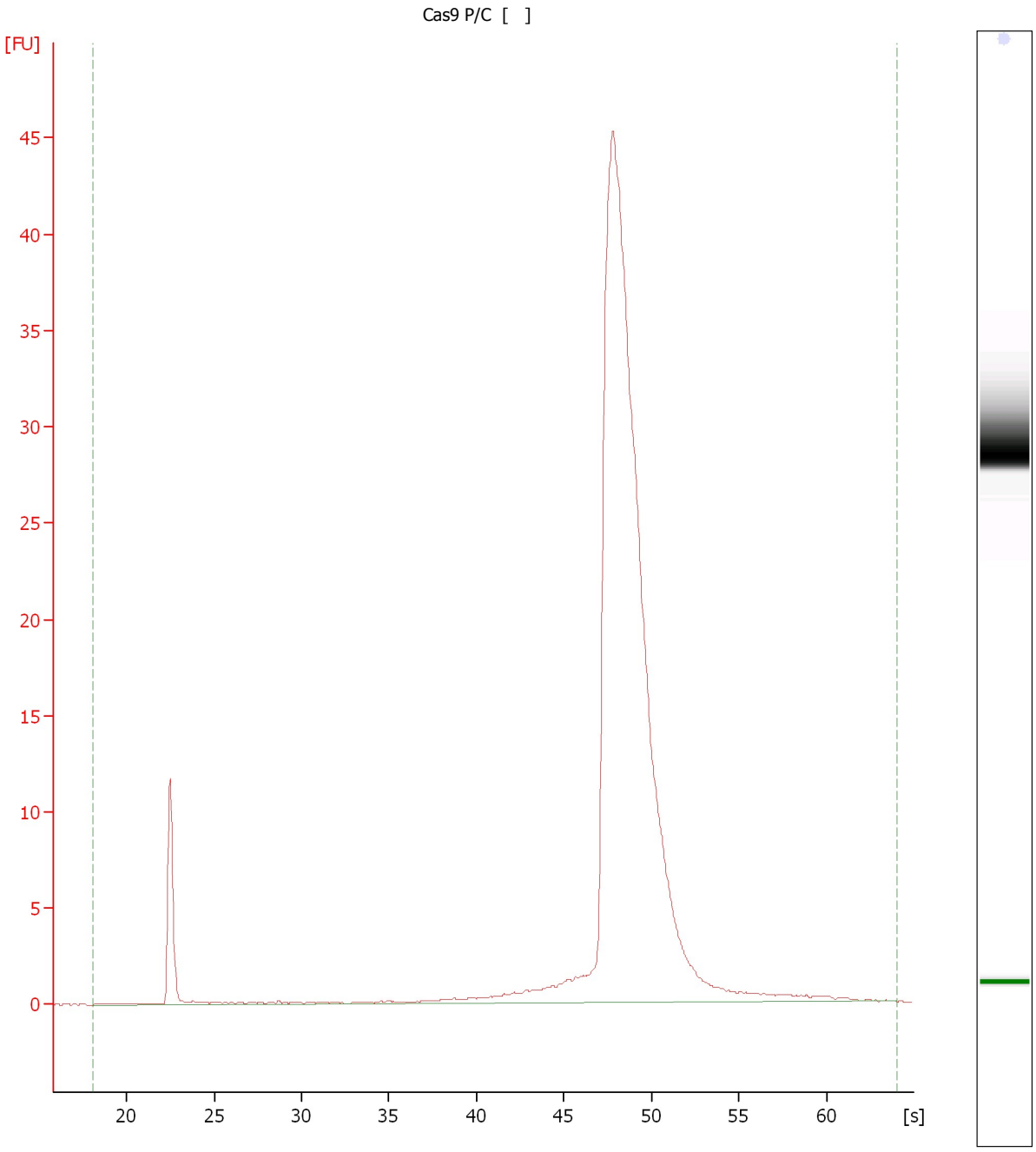


Cas9 mRNA

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-17_13-05-28.xad

Created: 17/01/2014 13:05:27
Modified: 17/01/2014 13:17:09

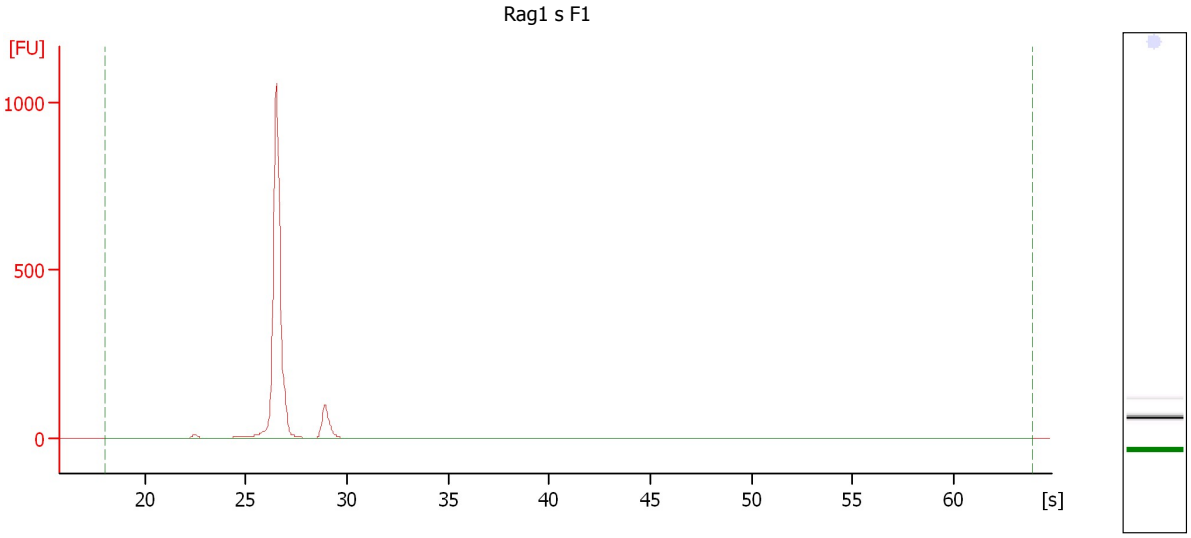
Electropherogram Summary



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-27_17-18-26.xad

Created: 27/01/2014 17:18:26
Modified: 27/01/2014 17:34:09

Electropherogram Summary Continued ...



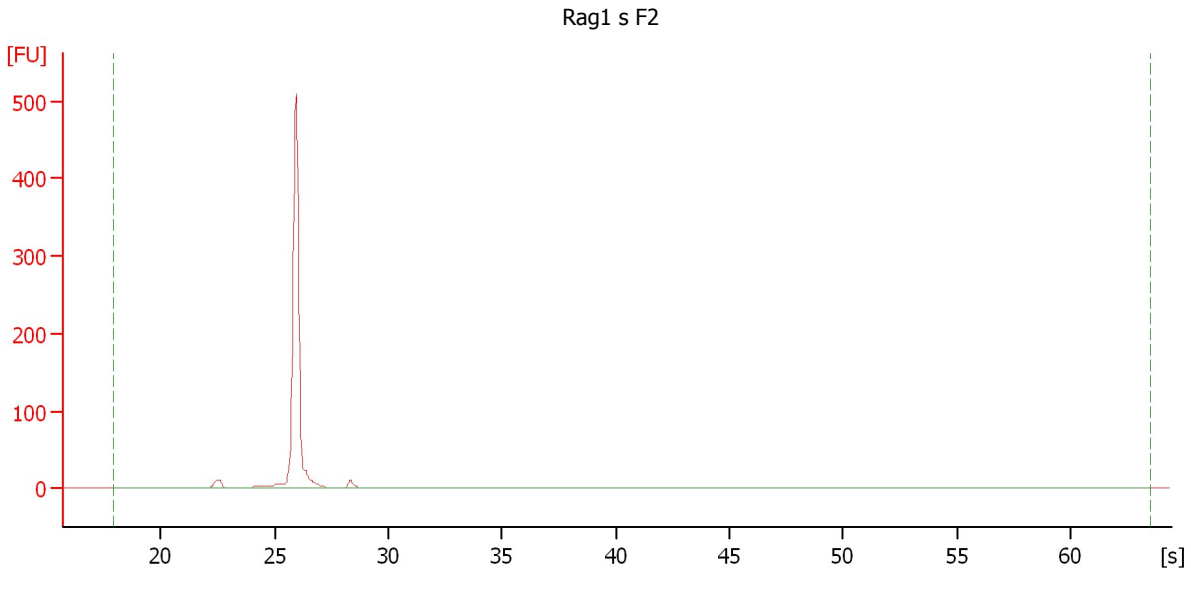
Overall Results for sample 1 : Rag1 s F1

RNA Area:	1,798.5	RNA Integrity Number (RIN):	2.6 (B.02.08)
RNA Concentration:	1,121 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.0	Result Flagging Label:	RIN: 2.60

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-27_17-18-26.xad

Created: 27/01/2014 17:18:26
Modified: 27/01/2014 17:34:09

Electropherogram Summary Continued ...



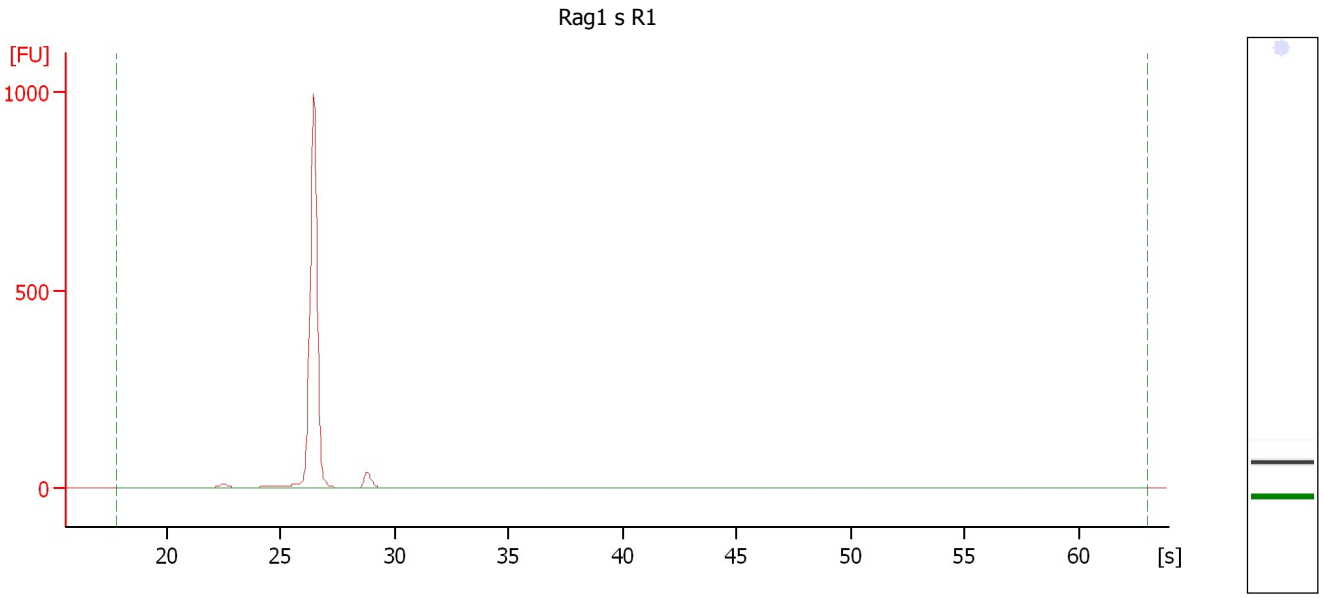
Overall Results for sample 2 : Rag1 s F2

RNA Area:	681.9	RNA Integrity Number (RIN):	2.6 (B.02.08)
RNA Concentration:	425 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.0	Result Flagging Label:	RIN: 2.60

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-27_17-18-26.xad

Created: 27/01/2014 17:18:26
Modified: 27/01/2014 17:34:09

Electropherogram Summary Continued ...



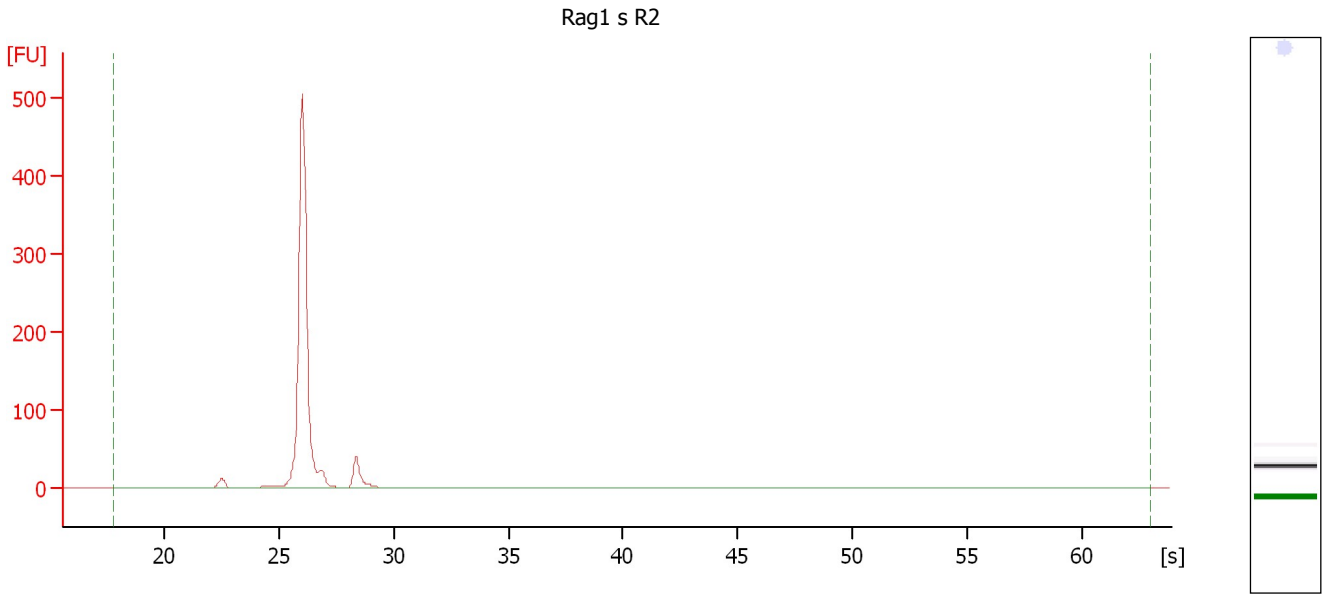
Overall Results for sample 3 : Rag1 s R1

RNA Area:	1,518.6	RNA Integrity Number (RIN):	2.6 (B.02.08)
RNA Concentration:	946 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.0	Result Flagging Label:	RIN: 2.60

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-27_17-18-26.xad

Created: 27/01/2014 17:18:26
Modified: 27/01/2014 17:34:09

Electropherogram Summary Continued ...



Overall Results for sample 4 : Rag1 s R2

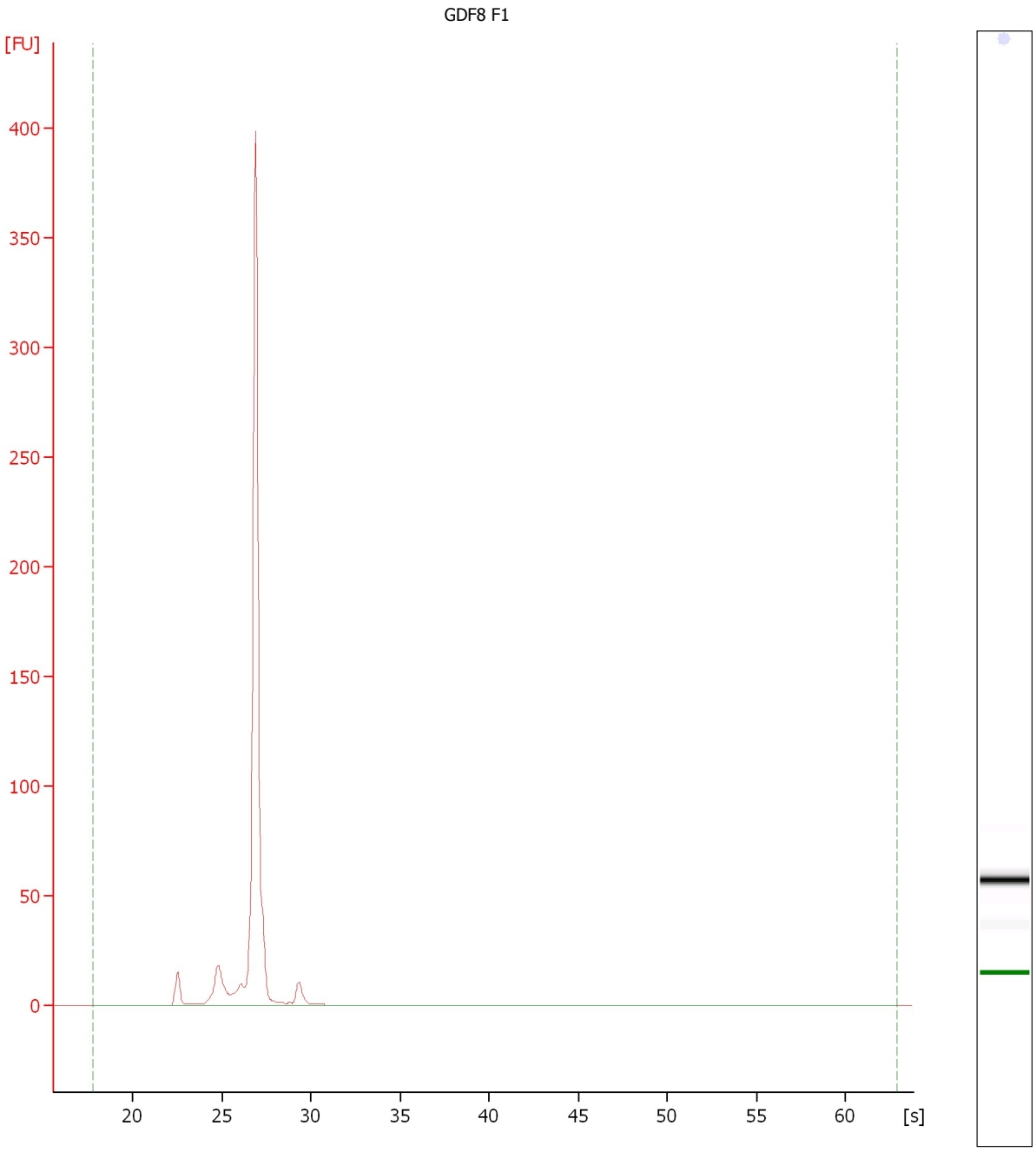
RNA Area:	883.0	RNA Integrity Number (RIN):	2.6 (B.02.08)
RNA Concentration:	550 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	0.0	Result Flagging Label:	RIN: 2.60

/subsection*MSTN sgRNAs

Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-16_16-57-43.xad

Created: 16/01/2014 16:57:42
Modified: 16/01/2014 17:18:48

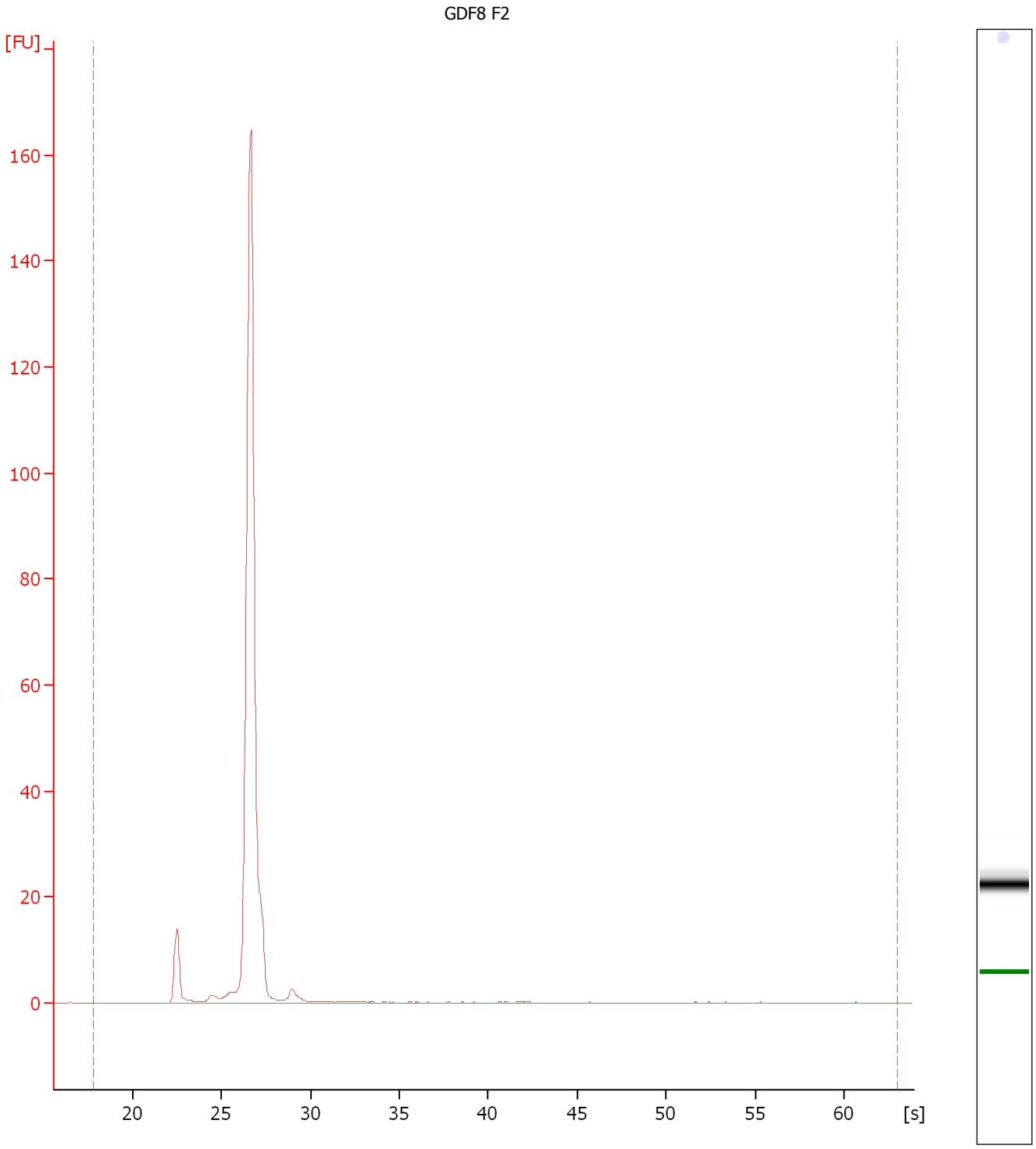
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-16_16-57-43.xad

Created: 16/01/2014 16:57:42
Modified: 16/01/2014 17:18:48

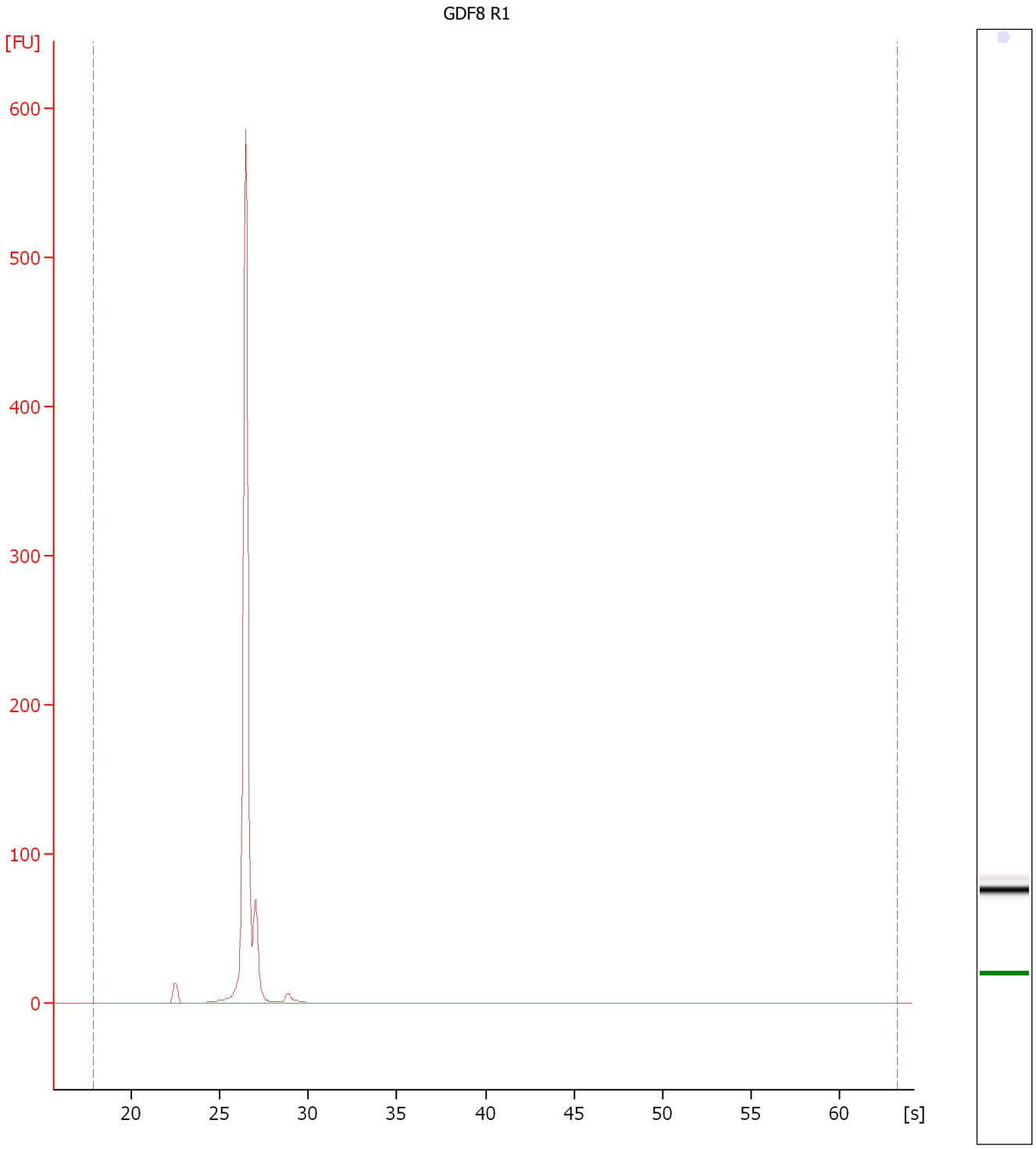
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-16_16-57-43.xad

Created: 16/01/2014 16:57:42
Modified: 16/01/2014 17:18:48

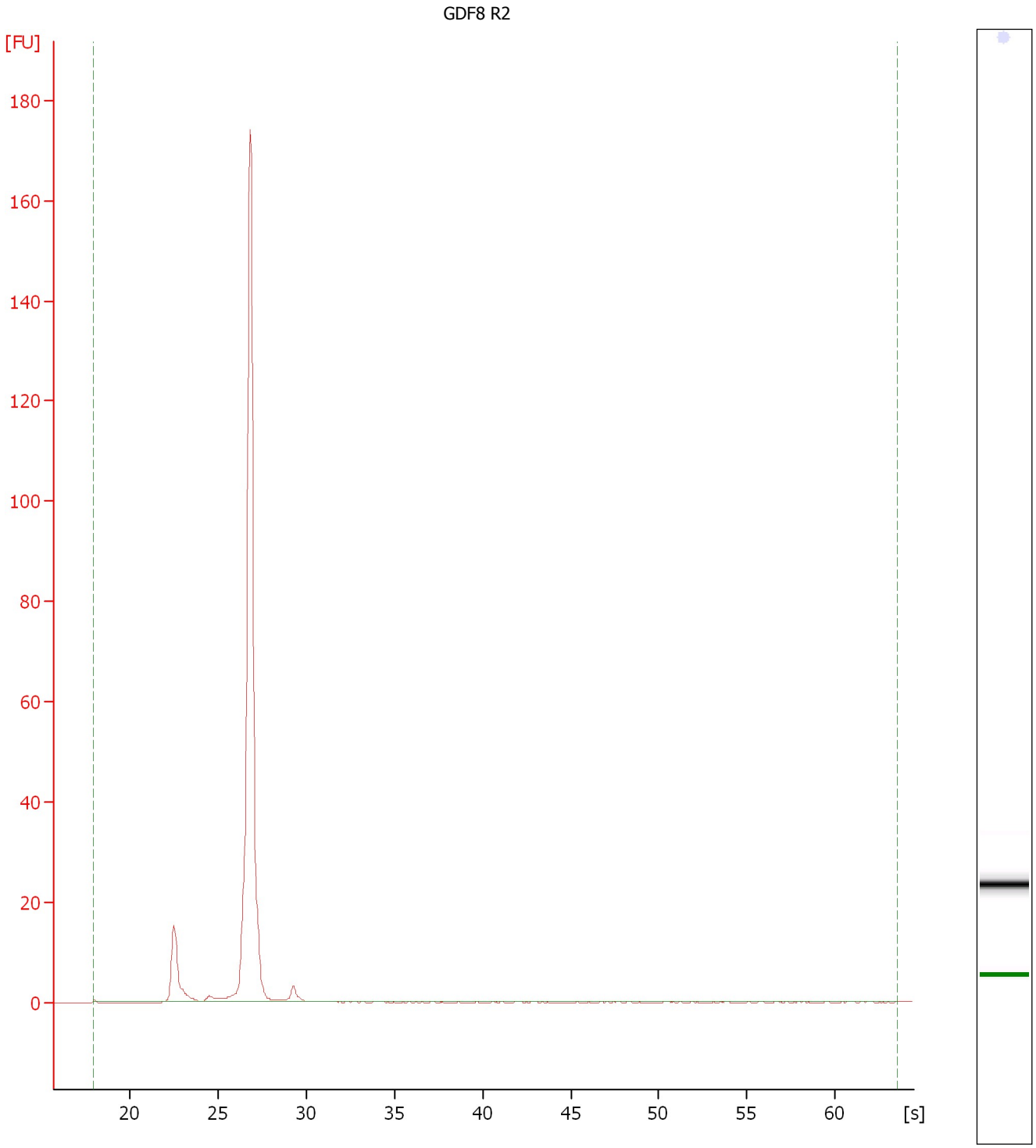
Electropherogram Summary Continued ...



Assay Class: Eukaryote Total RNA Nano
Data Path: C:\...Eukaryote Total RNA Nano_DE34903140_2014-01-16_16-57-43.xad

Created: 16/01/2014 16:57:42
Modified: 16/01/2014 17:18:48

Electropherogram Summary Continued ...



A.5. FMDV targets - breed mRNA alignments

On the following pages are alignments of the mRNA sequence of eIF4A1, eIF4G1 and IKBKG in different livestock species. In each figure, the bar across the top of each set of alignments indicates the degree of conservation between the individual sequences, with red showing that all sequences share the same nucleotides at that particular point. The top sequence in each alignment is the reference sequence from the genome of each species.

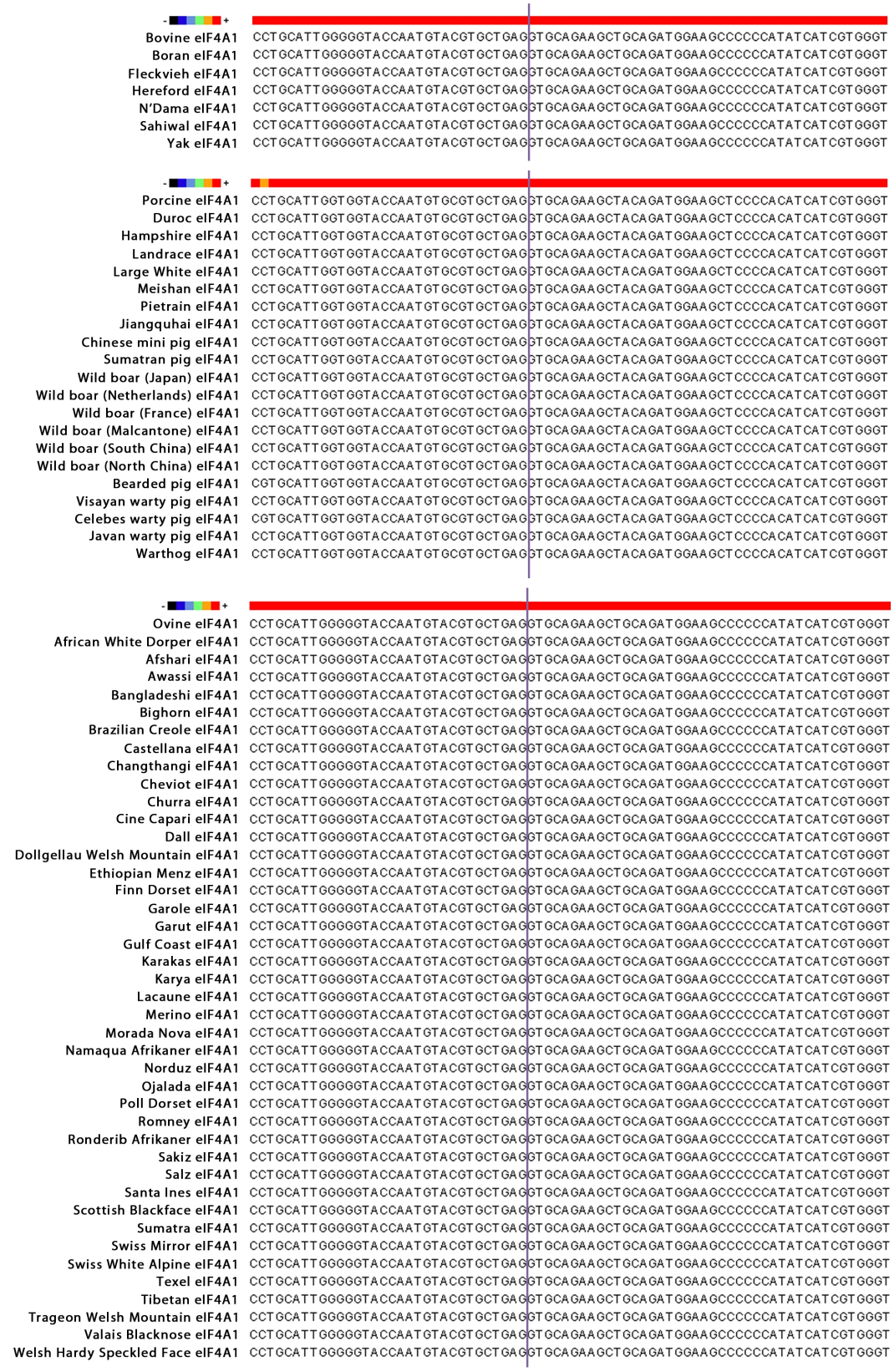


Figure A.1.: Nucleotide alignments of eIF4A1 mRNA sequences from various bovine, porcine and ovine breeds. The 3C cleavage site is indicated by the purple line.

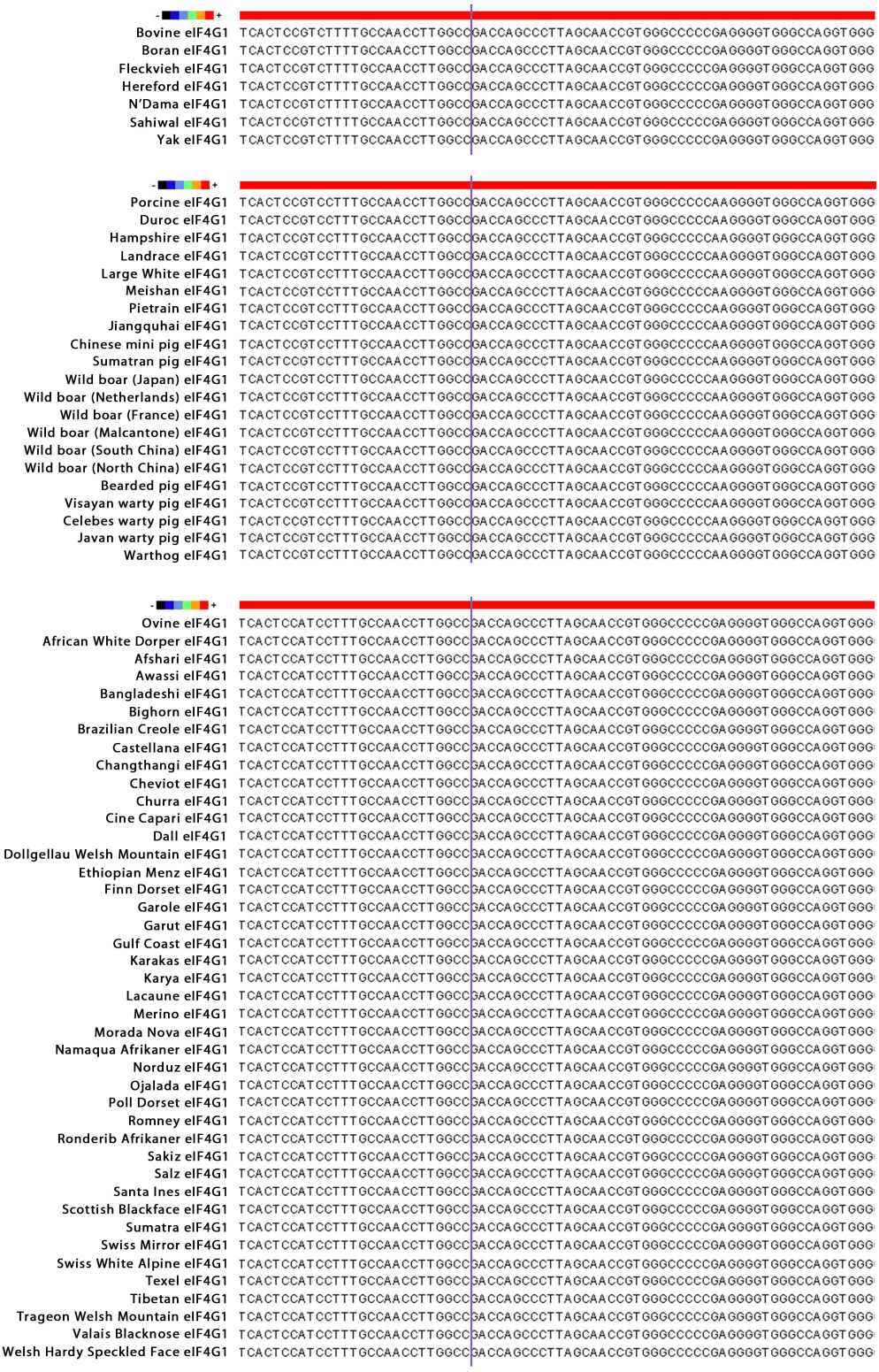


Figure A.2.: Nucleotide alignments of eIF4G1 mRNA sequences from various bovine, porcine and ovine breeds. The L protease cleavage site is indicated by the purple line

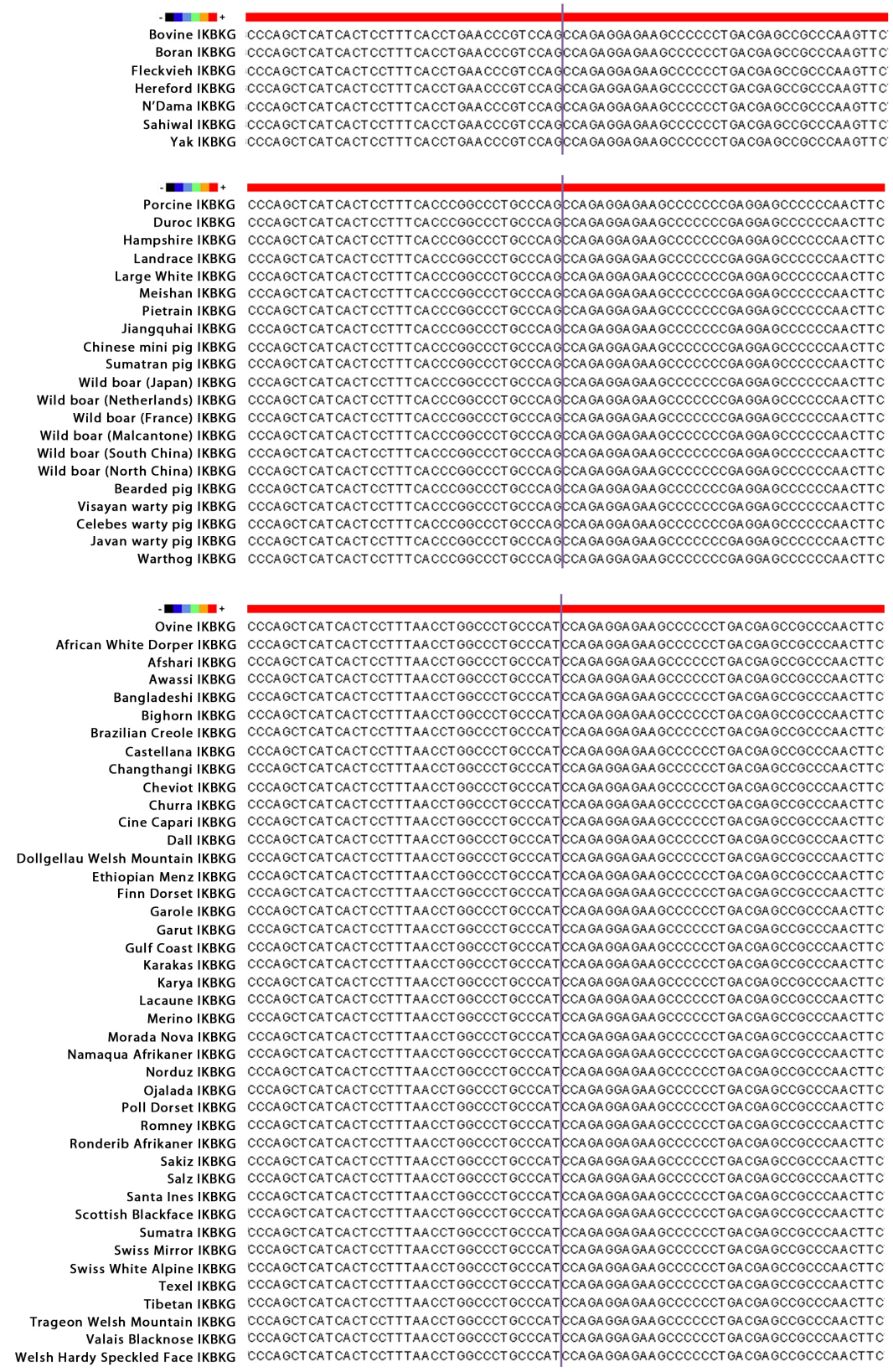


Figure A.3.: Nucleotide alignments of IKBKG mRNA sequences from various bovine, porcine and ovine breeds. The 3C protease cleavage site is indicated by the purple line

A.6. Predicted FMDV 3C targets

On the next pages are the complete lists of targets from the FMDV protease bioinfomatics work described in Section 5.3.3. As described previously, targets have been sorted firstly by cleavage score, then by surface score.Targets which did not meet the cleavage score cut off point of 0.8 are shaded in green.

A.6.1. Predicted bovine targets for FMDV 3C protease

Protein name	Uniprot identifiers	P1 amino acid	Cleavage score	Surface score	Predicted cleavage site	Likelihood of cleavage site
Collectin-43	P42916|CL43_BOVIN	E264	0.856	0.779	DISKEGKFT	Potential
TCF3 fusion partner homolog	Q17QH7|TFPT_BOVIN	E164	0.856	0.779	PPDKEGLSP	Potential
Oxysterol-binding protein (Fragment)	E1BA05|E1BA05_BOVIN	E363	0.856	0.764	EPLKETTYT	Potential
Protein DJ-1	Q5E946|PARK7_BOVIN	E64	0.855	0.792	DAKKEGPYD	Potential
Isoform V1 of Versican core protein	P81282-2|CSPG2_BOVIN	E1351	0.853	0.775	DARTEGPFT	Potential
Versican core protein	F1N6I5|F1N6I5_BOVIN	E1351	0.853	0.775	DARTEGPFT	Potential
Rap guanine nucleotide exchange factor 2	F1MSG6|RPGF2_BOVIN	E1390	0.848	0.775	IARKEGRYR	Potential
Protein KASH5	Q2T9R2|KASH5_BOVIN	E155	0.847	0.787	DPRPELPAT	Potential
Neuroendocrine convertase 2	Q9GLR0|NEC2_BOVIN	E64	0.845	0.787	LPFAEGLYH	Potential
DNA-directed RNA polymerase III subunit RPC8	F1ML03|F1ML03_BOVIN	E181	0.843	0.795	LPKKEAPYT	Potential
DNA-directed RNA polymerase III subunit RPC8	Q2T9X1|RPC8_BOVIN	E181	0.843	0.795	LPKKEAPYT	Potential
Mitochondrial import inner membrane translocase subunit Tim17-B	F1MS17|F1MS17_BOVIN	E158	0.841	0.78	LPSKEGTPG	Potential
Mitochondrial import inner membrane translocase subunit Tim17-B	Q2HJE9|TI17B_BOVIN	E158	0.841	0.78	LPSKEGTPG	Potential
Zinc finger with UFM1-specific peptidase domain protein	G5E563|G5E563_BOVIN	E404	0.838	0.77	DAWKEGFDP	Potential
Zinc finger with UFM1-specific peptidase domain protein	Q3SWY8|ZUFSP_BOVIN	E404	0.838	0.77	DAWKEGFDP	Potential
CDYL protein	A7MBK0|A7MBK0_BOVIN	E65	0.837	0.809	EKQKEGTFT	Potential
Chromodomain protein Y-like protein transcript variant 1	B2XBK5|B2XBK5_BOVIN	E88	0.837	0.809	EKQKEGTFT	Potential
Pulmonary surfactant-associated protein D	P35246|SFTPD_BOVIN	E315	0.837	0.776	DTRKEGTFI	Potential
Docking protein 2	A7MBB8|DOK2_BOVIN	E268	0.837	0.762	LPRPESPYA	Potential
Activating signal cointegrator 1 complex subunit 3	E1BNG3|ASCC3_BOVIN	E2017	0.835	0.759	MIEKELPAP	Potential
Neuroendocrine convertase 2	A7MBJ9|A7MBJ9_BOVIN	E64	0.833	0.795	LPFAEGLYQ	Potential
Fermitin family homolog 3	Q32LP0|URP2_BOVIN	E497	0.832	0.794	DASAEGLNP	Potential
Fermitin family homolog 3 (Fragment)	F1MMJ5|F1MMJ5_BOVIN	E107	0.832	0.794	DASAEGLNP	Potential
Solute carrier family 25 member 40	Q0VCH6|S2540_BOVIN	E6	0.831	0.78	DPESEGPAV	Potential
MOB kinase activator 3A	Q58D63|MOB3A_BOVIN	E210	0.831	0.769	EPLKEMTAR	Potential
KCNJ10 protein	A2VDQ4|A2VDQ4_BOVIN	E368	0.831	0.755	QAEKEGSAV	Potential
Steroid 17-alpha-hydroxylase/17,20 lyase	P05185|CP17A_BOVIN	E505	0.83	0.759	EAQAEGSTP	Potential
LMOD1 protein	A4IFK3|A4IFK3_BOVIN	E175	0.828	0.817	EAGKEGPAA	Potential
Cytochrome c oxidase subunit 6B1	P00429|CX6B1_BOVIN	E79	0.828	0.789	DRRAEGTFP	Potential
TAP binding protein (Tapasin)	Q2NL04|Q2NL04_BOVIN	E265	0.828	0.787	RPSQEGTYL	Potential
Ankyrin repeat and SOCS box protein 2	Q3SX45|ASB2_BOVIN	E135	0.828	0.753	EPNKEGWLP	Potential
High mobility group nucleosome-binding domain-containing protein 3	Q3ZBV4|HMGN3_BOVIN	E73	0.827	0.808	EAGKEGTAA	Potential

Coiled-coil domain-containing protein 86	F1MM20|F1MM20_BOVIN	E202	0.827	0.775	EPAREGPAP	Potential
Coiled-coil domain-containing protein 86	Q2TBX7|CCD86_BOVIN	E202	0.827	0.775	EPAREGPAP	Potential
Calcium-binding and coiled-coil domain-containing protein 1	F1N037|F1N037_BOVIN	E656	0.827	0.758	PPWKECPIC	Potential
Calcium-binding and coiled-coil domain-containing protein 1	Q2KJ21|CACO1_BOVIN	E656	0.827	0.758	PPWKECPIC	Potential
Cyclic nucleotide-gated cation channel beta-1	Q28181-1|CNGB1_BOVIN	E835	0.826	0.769	PPRPEPPAP	Potential
Isoform CNG4C of Cyclic nucleotide-gated cation channel beta-1	Q28181-2|CNGB1_BOVIN	E826	0.826	0.769	PPRPEPPAP	Potential
Isoform CNG4D of Cyclic nucleotide-gated cation channel beta-1	Q28181-3|CNGB1_BOVIN	E817	0.826	0.769	PPRPEPPAP	Potential
Isoform GARP2 of Cyclic nucleotide-gated cation channel beta-1	Q28181|CNGB1_BOVIN	E1291	0.826	0.769	PPRPEPPAP	Potential
DNAation factor, 45kDa, alpha polypeptide	Q0VC37|Q0VC37_BOVIN	E245	0.826	0.764	LALKEKPAP	Potential
BCL2/adenovirus E1B 19kD interacting protein like	Q17QP4|Q17QP4_BOVIN	E150	0.825	0.77	LPRAEGLGA	Potential
Visual system homeobox 1	Q9GMA3|VSX1_BOVIN	E299	0.825	0.752	DHLKEGSSP	Potential
Inactive rhomboid protein 1	A7YWH9|RHDF1_BOVIN	E329	0.825	0.75	RKQKEGGAA	Potential
Glutamate decarboxylase-like protein 1	A6QM00|GADL1_BOVIN	E262	0.824	0.784	EAKKEGAAP	Potential
TRIO protein	A6QQZ8|A6QQZ8_BOVIN	E907	0.824	0.767	CPAAEGWIP	Potential
HPN protein	A1L5C6|A1L5C6_BOVIN	E5	0.823	0.784	MAEKEGGRT	Potential
Enoyl-CoA hydratase, mitochondrial	Q58DM8|ECHM_BOVIN	E274	0.823	0.781	EDRKEGMAA	Potential
GGNBP1 protein	A6H760|A6H760_BOVIN	E227	0.823	0.775	FPLKEGLPR	Potential
Non-structural maintenance of chromosomes element 1 homolog	Q3T0X7|NSE1_BOVIN	E163	0.823	0.772	LIEKEGEFT	Potential
Non-structural maintenance of chromosomes element 1 homolog (Fragment)	F1MUB2|F1MUB2_BOVIN	E117	0.823	0.772	LIEKEGEFT	Potential
Regulator of microtubule dynamics protein 3	Q1JQC5|RMD3_BOVIN	E172	0.822	0.806	DAESEGGYT	Potential
V-type proton ATPase subunit F	Q28029|VATF_BOVIN	E95	0.822	0.754	IPSKEHPYD	Potential
Transporter (Fragment)	F6QN23|F6QN23_BOVIN	E619	0.821	0.766	TPLKEGLMA	Potential
Sepiapterin reductase	Q17QK8|SPRE_BOVIN	E96	0.82	0.785	LPRPEGLQR	Potential
Calpain-1 catalytic subunit	Q27970|CAN1_BOVIN	E508	0.819	0.773	EPNKEGDFV	Potential
DNA repair protein XRCC3	Q08DH8|XRCC3_BOVIN	E72	0.818	0.808	LLRQEGPFP	Potential
RAVER1 protein	A4FUZ2|A4FUZ2_BOVIN	E690	0.818	0.808	EPSPEGSYV	Potential
Semaphorin-3C	A7MB70|SEM3C_BOVIN	E357	0.818	0.777	FAHKEGPNH	Potential
RHBDD3 protein	A6H705|A6H705_BOVIN	E251	0.818	0.77	WPLSEGSAP	Potential
Calsyntenin-3	Q0VCN6-2|CSTN3_BOVIN	E104	0.818	0.767	EAQKEHTFT	Potential
Isoform 2 of Calsyntenin-3	Q0VCN6|CSTN3_BOVIN	E104	0.818	0.767	EAQKEHTFT	Potential
LOC508041 protein	A6QPL9|A6QPL9_BOVIN	E11	0.818	0.762	RCRKELTAA	Potential
CD3e molecule, epsilon associated protein	Q0VD10|Q0VD10_BOVIN	E417	0.818	0.757	EPQAEGMEP	Potential
Pre-miRNA 5'-monophosphate methyltransferase	Q29S19|BN3D2_BOVIN	E122	0.817	0.783	RAEKECPFP	Potential
Acetyl-CoA carboxylase 1	E1BGH6|E1BGH6_BOVIN	E1709	0.817	0.782	LARAEGIPR	Potential
Acetyl-CoA carboxylase 1	Q9TTS3|ACACA_BOVIN	E1709	0.817	0.782	LARAEGIPR	Potential
Acetolactate synthase-like protein	A6QQT9|ILVBL_BOVIN	E258	0.816	0.782	EPQPEGPLP	Potential
Acetolactate synthase-like protein (Fragment)	G1K1S1|G1K1S1_BOVIN	E265	0.816	0.782	EPQPEGPLP	Potential

Conglutinin	P23805|CONG_BOVIN	E315	0.816	0.772	DISTEGRFT	Potential
Centromere protein X	Q2NKU0|CENPX_BOVIN	E58	0.816	0.757	QAQAEGLAH	Potential
Probable RNA polymerase II nuclear localization protein SLC7A6OS	Q1JQE2|S7A6O_BOVIN	E101	0.815	0.779	EIRQEGRYR	Potential
ASAM protein	A5D7I8|A5D7I8_BOVIN	E106	0.815	0.775	KPSDEGRYT	Potential
Complement factor B	P81187|CFAB_BOVIN	E601	0.815	0.766	LPCTEGSIQ	Potential
PHD finger protein 20-like protein 1	A2VE56|P20L1_BOVIN	E113	0.814	0.828	AINKEGTFT	Potential
PHD finger protein 20-like protein 1	E1BGZ9|E1BGZ9_BOVIN	E113	0.814	0.828	AINKEGTFT	Potential
Ubiquitin carboxyl-terminal hydrolase 10	A5PJS6|UBP10_BOVIN	E717	0.814	0.751	EISKELLSP	Potential
Radial spoke head 10 homolog B	F1N5C9|F1N5C9_BOVIN	E600	0.813	0.766	MINKELTAT	Potential
Inactive ribonuclease-like protein 10	Q70IB2|RNS10_BOVIN	E58	0.812	0.807	EATKEGLAS	Potential
Protein FAM188B	A1A4L4|F188B_BOVIN	E245	0.812	0.778	QPHKEGSPQ	Potential
Protein FAM188B	F1MGQ9|F1MGQ9_BOVIN	E245	0.812	0.778	QPHKEGSPQ	Potential
Interleukin-12 receptor subunit beta-2	F1MEB5|F1MEB5_BOVIN	E812	0.812	0.76	LPSHEAPIT	Potential
Interleukin-12 receptor subunit beta-2	Q9BEG2|I12R2_BOVIN	E812	0.812	0.76	LPSHEAPIT	Potential
Isoform 2 of Microtubule-associated protein 4	P36225-3|MAP4_BOVIN	E368	0.811	0.816	LASAEGTVP	Potential
Isoform 3 of Microtubule-associated protein 4	P36225-4|MAP4_BOVIN	E368	0.811	0.816	LASAEGTVP	Potential
Isoform 4 of Microtubule-associated protein 4	P36225|MAP4_BOVIN	E368	0.811	0.816	LASAEGTVP	Potential
Microtubule-associated protein	F1MAZ1|F1MAZ1_BOVIN	E368	0.811	0.816	LASAEGTVP	Potential
Microtubule-associated protein	F1MAZ3|F1MAZ3_BOVIN	E368	0.811	0.816	LASAEGTVP	Potential
Microtubule-associated protein	F1N0J2|F1N0J2_BOVIN	E368	0.811	0.816	LASAEGTVP	Potential
Microtubule-associated protein	G3N2G7|G3N2G7_BOVIN	E368	0.811	0.816	LASAEGTVP	Potential
Microtubule-associated protein 4	P36225-2|MAP4_BOVIN	E368	0.811	0.816	LASAEGTVP	Potential
LMO7 protein	A8E4L8|A8E4L8_BOVIN	E1024	0.811	0.79	SPDKEGTRA	Potential
Protein CASC3	A5D7H5|CASC3_BOVIN	E429	0.811	0.781	PPPPEGLTP	Potential
FAP protein	A5D7B7|A5D7B7_BOVIN	E565	0.811	0.775	LASKEGIVI	Potential
DNA (cytosine-5)-methyltransferase 1	Q24K09|DNMT1_BOVIN	E448	0.81	0.794	DPSPEGGIN	Potential
Myocyte enhancer factor 2D	F1CYZ1|F1CYZ1_BOVIN	E492	0.81	0.793	EPEAEGSAV	Potential
Dihydropyrimidine dehydrogenase [NADP(+)]	F1N549|F1N549_BOVIN	E726	0.809	0.777	RAAKEGGAN	Potential
Dihydropyrimidine dehydrogenase [NADP(+)]	Q28007|DPYD_BOVIN	E726	0.809	0.777	RAAKEGGAN	Potential
NBR1 protein	A6QQS9|A6QQS9_BOVIN	E452	0.808	0.811	APTLEGTYT	Potential
Calcium binding tyrosine-(Y)-phosphorylation regulated	Q32L61|Q32L61_BOVIN	E435	0.808	0.777	PPVPEGLYE	Potential
Activating signal cointegrator 1 complex subunit 1	Q0IIJ9|Q0IIJ9_BOVIN	E292	0.807	0.804	DPNAEGRYN	Potential
LOC781700 protein	A5D7L4|A5D7L4_BOVIN	E225	0.807	0.796	SPSKEGRPA	Potential
Immediate early response 2	Q3T0B5|Q3T0B5_BOVIN	E166	0.806	0.781	PAQAEGAFP	Potential
Matrix-remodeling-associated protein 8	Q148M6|MXRA8_BOVIN	E283	0.806	0.779	EPADEGTYS	Potential
Protein-lysine methyltransferase METTL21C	A6QP81|MT21C_BOVIN	E16	0.806	0.775	RPLDEGPSP	Potential
Lysosomal-associated transmembrane protein 4A	Q6QRN8|LAP4A_BOVIN	E225	0.806	0.761	MPEKEPPPP	Potential

Delta-like protein (Fragment)	G3X6N5|G3X6N5_BOVIN	E965	0.806	0.75	LPREEGPNN	Potential
PTPRR protein	A5PKF8|A5PKF8_BOVIN	E226	0.805	0.783	IWSKEGFYA	Potential
SPRY domain-containing protein 7	Q2T9X3|SPRY7_BOVIN	E24	0.805	0.777	IPLKEMPAV	Potential
Cdc42 effector protein 1	Q17QW1|BORG5_BOVIN	E309	0.805	0.775	DARAEGLGA	Potential
Glutamate decarboxylase-like protein 1	F1MYA7|F1MYA7_BOVIN	E262	0.805	0.758	KAKKEGAAP	Potential
Pro-interleukin-16	Q0V8R5|IL16_BOVIN	E449	0.804	0.789	LASQEGTIQ	Potential
Zinc finger protein 512	A4FV61|ZN512_BOVIN	E146	0.804	0.762	RIRKEPPAY	Potential
MOB kinase activator 3B	Q29RK9|MOB3B_BOVIN	E209	0.804	0.755	EPLKEMTSR	Potential
AIM1L protein	A8E4N1|A8E4N1_BOVIN	E228	0.803	0.782	NPRAEGSPS	Potential
ULK2 protein	A6QQH7|A6QQH7_BOVIN	E434	0.803	0.778	PPSLEGLVT	Potential
Tuftelin-interacting protein 11	F1MR60|F1MR60_BOVIN	E309	0.803	0.754	PPGKEARAP	Potential
Tuftelin-interacting protein 11	Q29RR5|TFP11_BOVIN	E309	0.803	0.754	PPGKEARAP	Potential
Rho GTPase-activating protein 10	Q08DP6|RHG10_BOVIN	E386	0.802	0.785	IPRPEGSAQ	Potential
Reticulon	G3MZ35|G3MZ35_BOVIN	E224	0.802	0.779	QPKAEGICP	Potential
Reticulon-3	G3N1F7|G3N1F7_BOVIN	E128	0.802	0.779	QPKAEGICP	Potential
Alpha-aminoadipic semialdehyde dehydrogenase	E1BFG0|E1BFG0_BOVIN	E391	0.802	0.76	EAKKEGGTV	Potential
Alpha-aminoadipic semialdehyde dehydrogenase	Q2KJC9|AL7A1_BOVIN	E391	0.802	0.76	EAKKEGGTV	Potential
LMAN2 protein	A6QP36|A6QP36_BOVIN	E101	0.801	0.756	ERSKEGSIW	Potential
Bucentaur-2	A0JBZ9|A0JBZ9_BOVIN	E196	0.801	0.75	PPRSEGGQH	Potential
Ubiquitin carboxyl-terminal hydrolase	E1BGW4|E1BGW4_BOVIN	E855	0.801	0.75	ERRKEGRAE	Potential
Replication protein A2, 32kDa	Q2KI86|Q2KI86_BOVIN	E223	0.8	0.753	CPRPEGLNF	Potential
Mitogen-activated protein kinase kinase kinase 12	F1N791|F1N791_BOVIN	E679	0.8	0.75	TPPSEGSAP	Potential
Oral-facial-digital syndrome 1 isoform 1	F1MB20|F1MB20_BOVIN	E642	0.799	0.783	NASAEGSSP	Potential
V-type proton ATPase 116 kDa subunit a isoform 2	O97681|VPP2_BOVIN	E364	0.799	0.78	IPTKETPPT	Potential
MARVELD3 protein	A6H721|A6H721_BOVIN	E120	0.799	0.755	PLRKEGLGR	Potential
BTB (POZ) domain containing 12	F1N7Q0|F1N7Q0_BOVIN	E1680	0.799	0.754	PASQESTAT	Potential
Lysosome-associated membrane glycoprotein 1	Q05204|LAMP1_BOVIN	E307	0.798	0.816	PDAKEGSFT	Potential
RNA guanylyltransferase and 5'-phosphatase	Q2KHX7|Q2KHX7_BOVIN	E82	0.798	0.785	DIEKEGIKY	Potential
Transcription factor NF-E2 45 kDa subunit	Q5EAD3|NFE2_BOVIN	E49	0.798	0.771	EPSFEPPAP	Potential
Ras-related protein Rab-6B	A6QR46|RAB6B_BOVIN	E186	0.798	0.765	EKSKEGMID	Potential
CTF18, chromosome transmission fidelity factor 18 homolog	E1B8Q4|E1B8Q4_BOVIN	E30	0.798	0.758	LAELEGSAA	Potential
FRS3 protein	A6H792|A6H792_BOVIN	E217	0.797	0.805	QPLPEGRAP	Potential
MAP2K2 protein	Q17QH2|Q17QH2_BOVIN	E20	0.797	0.799	PAIAEGPSP	Potential
Zinc finger SWIM domain-containing protein 8	A7E305|ZSWM8_BOVIN	E693	0.797	0.759	TPRPEGKVP	Potential
Zinc finger SWIM domain-containing protein 8	E1BCX3|E1BCX3_BOVIN	E1115	0.797	0.759	TPRPEGKVP	Potential
E3 ubiquitin-protein ligase TRIM56	E1BD59|TRI56_BOVIN	E239	0.797	0.75	LAEKEALAR	Potential
E3 ubiquitin-protein ligase TRIM56	G5E5F1|G5E5F1_BOVIN	E239	0.797	0.75	LAEKEALAR	Potential
Transcription factor ETV6	Q0VC65|ETV6_BOVIN	E261	0.797	0.75	SPRQEGTRV	Potential
Transcription factor ETV6 (Fragment)	F1MQT2|F1MQT2_BOVIN	E206	0.797	0.75	SPRQEGTRV	Potential
Beta-casein	P02666|CASB_BOVIN	E123	0.796	0.801	PKHKEMPFP	Potential
Clathrin heavy chain (Fragment)	F1MPU0|F1MPU0_BOVIN	E1374	0.796	0.789	DAWKEGQFK	Potential

Clathrin heavy chain 1	P49951|CLH1_BOVIN	E1388	0.796	0.789	DAWKEGQFK	Potential
Type 2 phosphatidylinositol 4,5-bisphosphate 4-phosphatase	Q3SZ48|TM55A_BOVIN	E37	0.796	0.786	SPRAELPPP	Potential
CLEC11A protein	A5D7L1|A5D7L1_BOVIN	E249	0.796	0.77	DRRAEGLYL	Potential
Endonuclease/exonuclease/phosphatase family domain-containing protein 1	Q3MHJ7|EEPD1_BOVIN	E550	0.795	0.759	WTRKEGPRS	Potential
MGC159566 protein	A6QQ29|A6QQ29_BOVIN	E279	0.795	0.754	AALKEGSTW	Potential
PTPN5 protein	A6QQD4|A6QQD4_BOVIN	E23	0.794	0.811	LDQAEGPAA	Potential
Platelet-activating factor acetylhydrolase 2, cytoplasmic	P79106|PAFA2_BOVIN	E163	0.794	0.756	EALKEEWIP	Potential
DPYSL3 protein	A7MBI5|A7MBI5_BOVIN	E82	0.794	0.75	RPGAEGDTP	Potential
Heat shock factor protein 1	Q08DJ8|HSF1_BOVIN	E311	0.794	0.75	SPRAEGASP	Potential
Kinesin-like protein KIF2A	F1N034|F1N034_BOVIN	E50	0.793	0.789	EPSPETPPP	Potential
Kinesin-like protein KIF2A	G3N277|G3N277_BOVIN	E77	0.793	0.789	EPSPETPPP	Potential
Kinesin-like protein KIF2A	Q2NL05|KIF2A_BOVIN	E50	0.793	0.789	EPSPETPPP	Potential
Tyrosine-protein kinase receptor (Fragment)	F1MQH1|F1MQH1_BOVIN	E621	0.793	0.784	EAFMEGIFT	Potential
TCEB3 protein	A6QPU3|A6QPU3_BOVIN	E109	0.793	0.774	EAEIEGDYP	Potential
von Willebrand factor	F1N0R5|F1N0R5_BOVIN	E781	0.793	0.755	NPRAEGLEC	Potential
von Willebrand factor (Fragment)	P80012|VWF_BOVIN	E783	0.793	0.755	NPRAEGLEC	Potential
Keratin 10 (Epidermolytic hyperkeratosis; keratosis palmaris et plantaris)	A6QNZ7|A6QNZ7_BOVIN	E378	0.793	0.752	LAETEGRYC	Potential
Keratin, type I cytoskeletal 10	P06394|K1C10_BOVIN	E378	0.793	0.752	LAETEGRYC	Potential
Solute carrier family 27 (Fatty acid transporter), member 4	Q0VCQ2|Q0VCQ2_BOVIN	E604	0.793	0.752	ELQKEGFDP	Potential
Ribosomal RNA processing 1 homolog (S. cerevisiae)	Q148H9|Q148H9_BOVIN	E332	0.792	0.818	QALAEGLFP	Potential
Optic atrophy 3 protein homolog	Q05B66|OPA3_BOVIN	E181	0.792	0.763	TPRDEGPPD	Potential
Fibroblast growth factor-binding protein 1	Q9MZ06|FGFP1_BOVIN	E26	0.791	0.773	EAKKEGRNR	Potential
AP-3 complex subunit delta-1	Q865S1|AP3D1_BOVIN	E1034	0.791	0.765	LARPEGSSV	Potential
UNC5B protein	A5PJP9|A5PJP9_BOVIN	E595	0.79	0.816	LPLSEGTQT	Potential
CCCTC-binding factor (Zinc finger protein)	Q08DH9|Q08DH9_BOVIN	E145	0.79	0.77	EVSKEGLAE	Potential
SAMD14 protein	A6QL51|A6QL51_BOVIN	E132	0.79	0.756	TASHEGLAA	Potential
Annexin	F1MU06|F1MU06_BOVIN	E210	0.789	0.797	LKQAEGPST	Potential
Annexin A9	Q3ZC08|ANXA9_BOVIN	E210	0.789	0.797	LKQAEGPST	Potential
Radial spoke head 10 homolog B	Q1JPG1|RS10B_BOVIN	E600	0.789	0.79	IINKELTAT	Potential
T-complex protein 1 subunit beta	Q3ZBH0|TCPB_BOVIN	E482	0.789	0.782	LDMKEGTIG	Potential
Paraspeckle component 1	Q1LZD9|PSPC1_BOVIN	E57	0.789	0.781	RPADEGGFT	Potential
CD97 antigen	F1MCN3|F1MCN3_BOVIN	E144	0.789	0.751	CINTEGSYT	Potential
Fibulin-1	F1MYN5|F1MYN5_BOVIN	E332	0.789	0.751	CINTEGSYT	Potential
DnaJ homolog subfamily C member 3	F1N036|F1N036_BOVIN	E281	0.788	0.823	ELIKEGRYT	Potential
DnaJ homolog subfamily C member 3	Q27968|DNJC3_BOVIN	E281	0.788	0.823	ELIKEGRYT	Potential
Cathepsin K	Q5E968|CATK_BOVIN	E112	0.788	0.758	IPDWEGRAP	Potential
RING finger protein 112	E1BFZ5|E1BFZ5_BOVIN	E201	0.787	0.779	LASGEGSWP	Potential
RING finger protein 112	Q08DF2|RN112_BOVIN	E201	0.787	0.779	LASGEGSWP	Potential
NADH dehydrogenase [ubiquinone] iron-sulfur protein 5	Q02379|NDUS5_BOVIN	E89	0.787	0.769	KLIKEGKYT	Potential
LOC407171 protein	A5D7Q0|A5D7Q0_BOVIN	E164	0.787	0.767	LLSKEGAAH	Potential
Trafficking protein particle complex subunit 9	F1MTG7|F1MTG7_BOVIN	E1118	0.787	0.763	GPSKELPPS	Potential

Trafficking protein particle complex subunit 9	Q32PH0|TPPC9_BOVIN	E1118	0.787	0.763	GPSKELPPS	Potential
Protein arginine N-methyltransferase 7	A6QQV6|ANM7_BOVIN	E597	0.786	0.806	PIHAEGTIE	Potential
Protein arginine N-methyltransferase 7	F1MQX7|F1MQX7_BOVIN	E597	0.786	0.806	PIHAEGTIE	Potential
Angiotensin-converting enzyme (Fragment)	F1MQJ0|F1MQJ0_BOVIN	E1154	0.786	0.773	SARLEGPFV	Potential
LIM domain only protein 3	Q2KIA3|LMO3_BOVIN	E138	0.786	0.773	GLMKEGYAP	Potential
Gap junction alpha-4 protein	A4IFL1|CXA4_BOVIN	E271	0.786	0.769	LPMSEGPSS	Potential
Gap junction protein	G1K1Q0|G1K1Q0_BOVIN	E271	0.786	0.769	LPMSEGPSS	Potential
DDB1- and CUL4-associated factor 15	Q3SZD5|DCA15_BOVIN	E378	0.785	0.788	PPAAEAPAP	Potential
Microspherule protein 1	Q08E44|Q08E44_BOVIN	E381	0.785	0.775	DLSLEGPAW	Potential
Calcitonin gene-related peptide type 1 receptor	A6QP74|CALRL_BOVIN	E358	0.785	0.771	PWRPEGKIA	Potential
Calcitonin gene-related peptide type 1 receptor	G3N3V9|G3N3V9_BOVIN	E359	0.785	0.771	PWRPEGKIA	Potential
Calcitonin gene-related peptide type 1 receptor	G3X6P6|G3X6P6_BOVIN	E358	0.785	0.771	PWRPEGKIA	Potential
Peroxisomal carnitine O-octanoyltransferase	G5E5R3|G5E5R3_BOVIN	E119	0.785	0.77	WPPKEGTQL	Potential
Peroxisomal carnitine O-octanoyltransferase	O19094|OCTC_BOVIN	E118	0.785	0.77	WPPKEGTQL	Potential
A disintegrin and metalloproteinase with thrombospondin motifs 2	P79331|ATS2_BOVIN	E959	0.784	0.789	DARPEGRRA	Potential
Plastin-3	A7E3Q8|PLST_BOVIN	E101	0.784	0.779	INRKEGICA	Potential
Plastin-3	F1MSB7|F1MSB7_BOVIN	E101	0.784	0.779	INRKEGICA	Potential
AP2-associated protein kinase 1	F1MH24|AAK1_BOVIN	E318	0.784	0.767	LLKKECPIP	Potential
AP2-associated protein kinase 1 (Fragment)	H2XJE8|H2XJE8_BOVIN	E263	0.784	0.767	LLKKECPIP	Potential
Inositol polyphosphate 1-phosphatase	P21327|INPP_BOVIN	E256	0.784	0.76	NPSSEGSCR	Potential
E3 ubiquitin-protein ligase RNF34	Q5E9J6|RNF34_BOVIN	E56	0.783	0.816	PASAEGPNI	Potential
Tensin-4	F1MN94|F1MN94_BOVIN	E429	0.783	0.816	TASPEGPAR	Potential
Tensin-4	Q32PJ7|TENS4_BOVIN	E429	0.783	0.816	TASPEGPAR	Potential
Coronin	F1MFD6|F1MFD6_BOVIN	E666	0.783	0.811	EPLQEGPGP	Potential
Coronin-7	Q0V8F1|CORO7_BOVIN	E666	0.783	0.811	EPLQEGPGP	Potential
Autophagy-related protein 13	Q08DY8|ATG13_BOVIN	E270	0.783	0.757	VPGKEGGVP	Potential
Coatomer subunit gamma-2	A2VE21|COPG2_BOVIN	E462	0.782	0.821	LLGKEGPRT	Potential
EXOC4 protein	A6QLD1|A6QLD1_BOVIN	E797	0.782	0.766	PLAKEGNYA	Potential
Fibulin 5	Q2KJ89|Q2KJ89_BOVIN	E148	0.782	0.762	CINTEGGYT	Potential
Fibulin-5	Q5EA62|FBLN5_BOVIN	E148	0.782	0.762	CINTEGGYT	Potential
LOC510340 protein (Fragment)	A6QPC7|A6QPC7_BOVIN	E177	0.781	0.806	PASPEGLAE	Potential
Protein sprouty homolog 2	Q08E39|SPY2_BOVIN	E188	0.781	0.796	CKCKECTYP	Potential
Basal cell adhesion molecule	Q9MZ08|BCAM_BOVIN	E460	0.781	0.787	QPKAEGSWT	Potential
Matrix metalloproteinase-23	F1N5K7|F1N5K7_BOVIN	E368	0.781	0.784	NAINEGTYT	Potential
Matrix metalloproteinase-23	Q2TBM7|MMP23_BOVIN	E368	0.781	0.784	NAINEGTYT	Potential
ST3 beta-galactoside alpha-2,3-sialyltransferase 2	Q6H8M9|Q6H8M9_BOVIN	E63	0.781	0.772	RLSKEGLAG	Potential
Leucine rich repeat neuronal 5	Q58DI7|Q58DI7_BOVIN	E301	0.781	0.763	WARAEGLSS	Potential
Transmembrane 7 superfamily member 3	Q08DX3|Q08DX3_BOVIN	E553	0.781	0.752	LAQIEGLFQ	Potential
Mitochondrial uncoupling protein 2	Q3SZI5|UCP2_BOVIN	E264	0.78	0.777	MLQKEGPQA	Potential

Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)	A0JN77|A0JN77_BOVIN	E128	0.78	0.775	QARAEGIPV	Potential
Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)	Q58DK4|DHAK_BOVIN	E128	0.78	0.775	QARAEGIPV	Potential
Inhibin beta A chain	P07995|INHBA_BOVIN	E143	0.78	0.773	EISKEGSDL	Potential
Protein MMS22-like	E1BGH8|MMS22_BOVIN	E982	0.78	0.77	QQEKELPAP	Potential
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2	F1N437|F1N437_BOVIN	E62	0.78	0.767	RIRAEGLNC	Potential
Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2	Q58DM4|ALKB2_BOVIN	E62	0.78	0.767	RIRAEGLNC	Potential
Fibrinogen alpha chain	A5PJE3|A5PJE3_BOVIN	E52	0.78	0.758	SACKETGWP	Potential
Fibrinogen alpha chain	P02672|FIBA_BOVIN	E52	0.78	0.758	SACKETGWP	Potential
C2 calcium-dependent domain-containing protein 4A	Q2KJ18|C2C4A_BOVIN	E279	0.779	0.791	LLRAEGPAG	Potential
C2 calcium-dependent domain-containing protein 4A (Fragment)	G3N1Z1|G3N1Z1_BOVIN	E268	0.779	0.791	LLRAEGPAG	Potential
Prenylated Rab acceptor protein 1	Q1RMH4|PRAF1_BOVIN	E14	0.779	0.79	DAEAEGLSA	Potential
Polymerase (DNA-directed), delta interacting protein 3	A7YW33|A7YW33_BOVIN	E266	0.779	0.783	EPPKELPPA	Potential
Serine beta-lactamase-like protein LACTB, mitochondrial	P83095|LACTB_BOVIN	E464	0.779	0.783	SWDKEGKYA	Potential
Histidine ammonia-lyase	A7YWP4|HUTH_BOVIN	E301	0.779	0.764	LKPKEGLAL	Potential
Fibronectin	P07589|FINC_BOVIN	E657	0.778	0.763	DRWKEATIP	Potential
S1PR3 protein	A6QR17|A6QR17_BOVIN	E304	0.778	0.755	LASKEMRRA	Potential
Transcription factor E3	E1BHE4|E1BHE4_BOVIN	E17	0.777	0.828	EASAEGPRA	Potential
Transcription factor E3	Q05B92|TFE3_BOVIN	E17	0.777	0.828	EASAEGPRA	Potential
LOC539976 protein	A4FUY3|A4FUY3_BOVIN	E247	0.777	0.806	RALCEGPYD	Potential
Ameloblastin	Q9XSX7|AMBN_BOVIN	E302	0.777	0.789	LANPEGNIP	Potential
Tetratricopeptide repeat protein 30A	A2VE45|TT30A_BOVIN	E166	0.777	0.75	LLYKEGHYE	Potential
Tetratricopeptide repeat protein 30B	A6H739|TT30B_BOVIN	E166	0.777	0.75	LLYKEGHYE	Potential
Hydroxysteroid (17-beta) dehydrogenase 1	Q1JQD0|Q1JQD0_BOVIN	E105	0.776	0.787	EAHKEGSVD	Potential
EIF2AK3 protein	A5D791|A5D791_BOVIN	E1015	0.776	0.779	LILFELLYP	Potential
CC2D1A protein	A5PKI6|A5PKI6_BOVIN	E934	0.776	0.768	EAAKEALYR	Potential
IQ domain-containing protein C	F1N4P6|F1N4P6_BOVIN	E51	0.776	0.768	LQWTEGWIP	Potential
Isoform 2 of IQ domain-containing protein C	Q2TBI7|IQCC_BOVIN	E51	0.776	0.768	LQWTEGWIP	Potential
Leucine zipper transcription factor-like protein 1	Q3ZBL4|LZTL1_BOVIN	E143	0.776	0.759	APLHEGGAA	Potential
Leucine zipper transcription factor-like protein 1 (Fragment)	F1MNG7|F1MNG7_BOVIN	E101	0.776	0.759	APLHEGGAA	Potential
TTLL11 protein	A7E2Z5|A7E2Z5_BOVIN	E273	0.776	0.755	LASKENPPH	Potential
Oxidation resistance protein 1	A5PKL1|OXR1_BOVIN	E424	0.776	0.753	MPTKEGDQA	Potential
Oxidation resistance protein 1 (Fragment)	F1MR28|F1MR28_BOVIN	E323	0.776	0.753	MPTKEGDQA	Potential
Thrombomodulin	F1N6M2|F1N6M2_BOVIN	E388	0.775	0.81	CICAEGFAP	Potential
Thrombomodulin (Fragment)	P06579|TRBM_BOVIN	E166	0.775	0.81	CICAEGFAP	Potential
PDK3 protein	A6QLG3|A6QLG3_BOVIN	E265	0.775	0.803	EDRKEGYPA	Potential

Galectin	G5E5J2|G5E5J2_BOVIN	E192	0.775	0.8	LPTMEGPPA	Potential
Galectin-4	Q3T0D6|LEG4_BOVIN	E192	0.775	0.8	LPTMEGPPA	Potential
Targeting protein for Xklp2	A6H6Z7|TPX2_BOVIN	E435	0.775	0.771	PEKKELPIT	Potential
Targeting protein for Xklp2	F1MJB6|F1MJB6_BOVIN	E435	0.775	0.771	PEKKELPIT	Potential
Pyruvate dehydrogenase protein X component	P22439|ODPX_BOVIN	E272	0.775	0.768	QPNVEGTFT	Potential
Cholesterol side-chain cleavage enzyme, mitochondrial	P00189|CP11A_BOVIN	E114	0.775	0.754	LFKFEGSYP	Potential
Cholesterol side-chain cleavage enzyme, mitochondrial (Fragment)	F1N060|F1N060_BOVIN	E25	0.775	0.754	LFKFEGSYP	Potential
Matrix metalloproteinase-9	F1MF56|F1MF56_BOVIN	E626	0.775	0.753	LPRPEGKVL	Potential
Matrix metalloproteinase-9	P52176|MMP9_BOVIN	E626	0.775	0.753	LPRPEGKVL	Potential
Putative UDP-galactose translocator	Q8SPM1|Q8SPM1_BOVIN	E254	0.774	0.824	LWWAEGTAV	Potential
UDP-galactose translocator	F1MZI7|F1MZI7_BOVIN	E254	0.774	0.824	LWWAEGTAV	Potential
UDP-galactose translocator	Q58DA6|S35A2_BOVIN	E254	0.774	0.824	LWWAEGTAV	Potential
Neurotrimin	Q58DA5|NTRI_BOVIN	E110	0.774	0.77	DVYDEGPYT	Potential
Opioid-binding protein/cell adhesion molecule	G5E5Q3|G5E5Q3_BOVIN	E110	0.774	0.77	DVYDEGPYT	Potential
Opioid-binding protein/cell adhesion molecule	P11834|OPCM_BOVIN	E110	0.774	0.77	DVYDEGPYT	Potential
Arrestin domain containing 2	A2VDR7|A2VDR7_BOVIN	E370	0.774	0.762	DMSMEGPFF	Potential
SF4 protein	A4FV64|A4FV64_BOVIN	E536	0.773	0.777	KALKEGREP	Potential
Cadherin-13	Q3B7N0|CAD13_BOVIN	E247	0.773	0.774	PIFREGPYI	Potential
Cadherin-13 (Fragment)	F1MKP6|F1MKP6_BOVIN	E197	0.773	0.774	PIFREGPYI	Potential
GA-binding protein subunit beta-2	Q0V8G2-2|GABP2_BOVIN	E320	0.773	0.772	EESKEGTER	Potential
Isoform 2 of GA-binding protein subunit beta-2	Q0V8G2|GABP2_BOVIN	E349	0.773	0.772	EESKEGTER	Potential
Trafficking protein particle complex subunit 11	A6QLC7|TPC11_BOVIN	E486	0.773	0.766	DYRSEGWWT	Potential
Trafficking protein particle complex subunit 11	F1MXL8|F1MXL8_BOVIN	E486	0.773	0.766	DYRSEGWWT	Potential
60S ribosomal protein L23a	Q24JY1|RL23A_BOVIN	E8	0.773	0.765	KAKKEAPAP	Potential
Carboxypeptidase Q	Q17QK3|CBPQ_BOVIN	E167	0.773	0.758	GPDAEGKIV	Potential
SEC14L1 protein	A7MBE2|A7MBE2_BOVIN	E616	0.772	0.757	LICKEGESV	Potential
Dispatched homolog 1 (Drosophila)	Q2T9Q5|Q2T9Q5_BOVIN	E44	0.771	0.833	LACKEAPRT	Potential
Protein arginine N-methyltransferase 5	A7YW45|ANM5_BOVIN	E499	0.771	0.776	EAQFEMPYV	Potential
Protein arginine N-methyltransferase 5	F1N443|F1N443_BOVIN	E499	0.771	0.776	EAQFEMPYV	Potential
Angiomotin-like protein 1	E1BEQ5|AMOL1_BOVIN	E233	0.771	0.76	KARTEGRPT	Potential
Cytosolic carboxypeptidase 3	E1B9D8|CBPC3_BOVIN	E515	0.771	0.753	QKSKEGTGR	Potential
Cytosolic carboxypeptidase 3	G3N121|G3N121_BOVIN	E515	0.771	0.753	QKSKEGTGR	Potential
Cytosolic carboxypeptidase 3	G5E5V2|G5E5V2_BOVIN	E520	0.771	0.753	QKSKEGTGR	Potential
Olfactory receptor	G5E5M2|G5E5M2_BOVIN	E100	0.771	0.752	LISFEGCMT	Potential
Polynucleotide 5'-hydroxyl-kinase NOL9	E1BPN0|NOL9_BOVIN	E478	0.77	0.772	DEEKEGPVM	Potential
Cytochrome c oxidase subunit 1	P00396|COX1_BOVIN	E40	0.77	0.759	LIRAELGQP	Potential
Cytochrome c oxidase subunit 1	Q6QTG9|Q6QTG9_BOVIN	E40	0.77	0.759	LIRAELGQP	Potential
Collagen alpha-2(XI) chain	F1MRP6|F1MRP6_BOVIN	E696	0.77	0.753	HPGKEGPPG	Potential
Collagen alpha-2(XI) chain	Q32S24|COBA2_BOVIN	E696	0.77	0.753	HPGKEGPPG	Potential
Melanoma-associated antigen D4	A6QLI5|MAGD4_BOVIN	E204	0.769	0.834	EAAAEGPST	Potential
REEP3 protein	A6QQT6|A6QQT6_BOVIN	E206	0.769	0.814	DEEAEGPYS	Potential

LRRTM2 protein	A6H701|A6H701_BOVIN	E492	0.769	0.808	EPTHEGPFI	Potential
General transcription factor IIE subunit 1	A6QLI8|T2EA_BOVIN	E297	0.769	0.77	EEMKEGGID	Potential
General transcription factor IIE subunit 1	F1MGT5|F1MGT5_BOVIN	E298	0.769	0.77	EEMKEGGID	Potential
Endothelin converting enzyme	Q28868|Q28868_BOVIN	E713	0.769	0.76	ESSHEGLIT	Potential
Endothelin-converting enzyme 1	P42891|ECE1_BOVIN	E709	0.769	0.76	ESSHEGLIT	Potential
Nuclear receptor subfamily 1 group D member 1	Q08E02|NR1D1_BOVIN	E397	0.769	0.753	YPAPEGEAP	Potential
TNFRSF6B protein	A6QPW7|A6QPW7_BOVIN	E244	0.768	0.751	PPLREGRAA	Potential
Collagen alpha-1(IV) chain	G1K238|G1K238_BOVIN	E1365	0.767	0.804	LPGPEGPAG	Potential
Aggrecan core protein	F1N367|F1N367_BOVIN	E1334	0.767	0.774	LPSGEGPEV	Potential
Aggrecan core protein	F1N368|F1N368_BOVIN	E1334	0.767	0.774	LPSGEGPEV	Potential
Aggrecan core protein	P13608-2|PGCA_BOVIN	E1333	0.767	0.774	LPSGEGPEV	Potential
Isoform 2 of Aggrecan core protein	P13608|PGCA_BOVIN	E1333	0.767	0.774	LPSGEGPEV	Potential
RSPRY1 protein	A4IFG8|A4IFG8_BOVIN	E424	0.767	0.769	PCWKEGDTV	Potential
Cyclic AMP-responsive element-binding protein 3	Q8SQ19|CREB3_BOVIN	E136	0.767	0.766	LLEKEGLTL	Potential
Importin subunit alpha	Q3SYV6|Q3SYV6_BOVIN	E354	0.767	0.764	NIQKEATWT	Potential
ETS1 protein	A5PJG9|A5PJG9_BOVIN	E151	0.767	0.76	PPYPESRYT	Potential
GPR182 protein (Fragment)	A6QPN2|A6QPN2_BOVIN	E403	0.766	0.8	IITKEGIQP	Potential
Polo-like kinase 3 (Drosophila)	Q0VCE7|Q0VCE7_BOVIN	E389	0.766	0.787	DARPEAPAA	Potential
Vesicular, overexpressed in cancer, prosurvival protein 1	A6QNZ8|VOPP1_BOVIN	E30	0.766	0.769	CWYFEGLYP	Potential
Cell division cycle protein 27 homolog	A7Z061|CDC27_BOVIN	E789	0.766	0.754	LPDDEEPIT	Potential
Cyclic GMP-AMP synthase	E1BGN7|CGAS_BOVIN	E84	0.765	0.823	EDQAEGPAA	Potential
ABI3 protein	A4FUD5|A4FUD5_BOVIN	E262	0.765	0.778	PPAPELPMP	Potential
Protein LTV1 homolog	Q0VC06|LTV1_BOVIN	E198	0.765	0.77	DDEKEGGSD	Potential
RECQL4 protein	A5D786|A5D786_BOVIN	E877	0.765	0.755	PPEQEGSRS	Potential
Insulin-like 3	E1BAS5|E1BAS5_BOVIN	E210	0.764	0.797	LIAQEGPFF	Potential
Chaperone activity of bc1 complex-like, mitochondrial	Q29RI0|ADCK3_BOVIN	E113	0.763	0.811	AAGSEGPAP	Potential
Matrix metallopeptidase 19	Q08DI9|Q08DI9_BOVIN	E199	0.763	0.797	ELWTEGTYR	Potential
Putative sodium-coupled neutral amino acid transporter 7	A7E3U5|S38A7_BOVIN	E229	0.763	0.786	WPDKEMTPA	Potential
Zinc finger FYVE domain-containing protein 21	Q05B78|ZFY21_BOVIN	E195	0.763	0.78	RPAAEGGNA	Potential
Zinc finger FYVE domain-containing protein 21 (Fragment)	F1MHW4|F1MHW4_BOVIN	E149	0.763	0.78	RPAAEGGNA	Potential
SALL2 protein	A6QQL8|A6QQL8_BOVIN	E82	0.763	0.766	EPRPEGHSS	Potential
Citrate synthase, mitochondrial	Q29RK1|CISY_BOVIN	E117	0.762	0.828	EPLPEGLFW	Potential
Serpin peptidase inhibitor, clade A (Alpha-1 antiproteinase, antitrypsin), member 7	Q3SYR0|Q3SYR0_BOVIN	E22	0.762	0.767	PNSCEGKIT	Potential
Thyroxine-binding globulin	Q9TT36|THBG_BOVIN	E22	0.762	0.767	PNSCEGKIT	Potential
Ribonucleoside-diphosphate reductase	F1MSF1|F1MSF1_BOVIN	E229	0.762	0.762	DDSIEGIYD	Potential
Collagen alpha-2(I) chain	P02465|CO1A2_BOVIN	E814	0.762	0.758	PAGKEGLRG	Potential
Coatomer subunit beta'	P35605|COPB2_BOVIN	E743	0.761	0.757	DACLELLIR	Potential
Cystic fibrosis transmembrane conductance regulator	F1MN01|F1MN01_BOVIN	E620	0.76	0.777	LILHEGSIY	Potential
Cystic fibrosis transmembrane conductance regulator	P35071|CFTR_BOVIN	E620	0.76	0.777	LILHEGSIY	Potential

Rho GTPase-activating protein 36	A7MB27|RHG36_BOVIN	E180	0.76	0.762	LAELEGGAL	Potential
SNF8, ESCRT-II complex subunit, homolog (S. cerevisiae)	Q08DR7|Q08DR7_BOVIN	E219	0.76	0.755	HLLKEGLAW	Potential
Protein S100-A12	P79105|S10AC_BOVIN	E40	0.759	0.795	LITKELPKT	Potential
Protein S100-A12 (Fragment)	G3N2H5|G3N2H5_BOVIN	E43	0.759	0.795	LITKELPKT	Potential
Cyclic nucleotide-gated cation channel beta-1	F1MTS6|F1MTS6_BOVIN	E840	0.759	0.782	PPAPEPPAP	Potential
Cyclic nucleotide-gated cation channel beta-1	F1MTS7|F1MTS7_BOVIN	E1282	0.759	0.782	PPAPEPPAP	Potential
Cyclic nucleotide-gated cation channel beta-1	F1N3J4|F1N3J4_BOVIN	E826	0.759	0.782	PPAPEPPAP	Potential
Cyclic nucleotide-gated cation channel beta-1	G3MZY3|G3MZY3_BOVIN	E836	0.759	0.782	PPAPEPPAP	Potential
Cyclic nucleotide-gated cation channel beta-1	G3N2E6|G3N2E6_BOVIN	E817	0.759	0.782	PPAPEPPAP	Potential
Actin-related protein 2/3 complex subunit 1B	Q58CQ2|ARC1B_BOVIN	E312	0.759	0.761	KASSEGSAA	Potential
Rab-like protein 6	Q08DA0|RABL6_BOVIN	E512	0.759	0.76	KPRPEGPSG	Potential
Drebrin 1	Q1RMT6|Q1RMT6_BOVIN	E531	0.759	0.758	LPAGEGSAN	Potential
Transporter	E1BHC0|E1BHC0_BOVIN	E561	0.759	0.755	PLRQEKPYP	Potential
LIM and senescent cell antigen-like-containing domain protein 2	Q2KJ33|LIMS2_BOVIN	E52	0.758	0.828	RPFPEGLFY	Potential
Asparagine synthetase [glutamine-hydrolyzing]	Q1LZA3|ASNS_BOVIN	E514	0.758	0.813	PKTKEGYYY	Potential
Cullin-associated NEDD8-dissociated protein 1	A7MBJ5|CAND1_BOVIN	E813	0.758	0.808	ACPKEGPAV	Potential
Fatty acid synthase	F1N647|F1N647_BOVIN	E1340	0.758	0.801	AALKEGGFL	Potential
Fatty acid synthase	Q71SP7|FAS_BOVIN	E1340	0.758	0.801	AALKEGGFL	Potential
Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe	Q2KJ92|LFNG_BOVIN	E109	0.758	0.766	RPLAEPLAP	Potential
Alpha-2-macroglobulin	Q7SIH1|A2MG_BOVIN	E612	0.758	0.752	RPEAELSAA	Potential
MARCKS-related protein	G3MY11|G3MY11_BOVIN	E119	0.758	0.751	RNRKEGGGD	Potential
MARCKS-related protein	G3MZA0|G3MZA0_BOVIN	E111	0.758	0.751	RNRKEGGGD	Potential
MARCKS-related protein	Q0VBZ9|MRP_BOVIN	E111	0.758	0.751	RNRKEGGGD	Potential
FTS and Hook-interacting protein	A7YY62|F16A2_BOVIN	E838	0.757	0.8	LDWTEGPAA	Potential
FTS and Hook-interacting protein (Fragment)	F1MLL5|F1MLL5_BOVIN	E730	0.757	0.8	LDWTEGPAA	Potential
Solute carrier family 3 (Cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino acid transport), member 1	Q3SZF7|Q3SZF7_BOVIN	E644	0.757	0.777	LASGEGLIL	Potential
MGC148449 protein	A4IFQ5|A4IFQ5_BOVIN	E16	0.757	0.775	LPGKEDPPA	Potential
Keratin, type II cuticular Hb3	A4FUZ0|KRT83_BOVIN	E161	0.757	0.773	EPLFEGYIE	Potential
Short transient receptor potential channel 5	F1MTE6|F1MTE6_BOVIN	E863	0.757	0.758	EPSSEPMYT	Potential
Zinc finger C2HC domain-containing protein 1B	Q32KN7|ZC21B_BOVIN	E102	0.756	0.786	LAIKEGRPL	Potential
Stress-70 protein, mitochondrial	Q3ZCH0|GRP75_BOVIN	E352	0.756	0.771	RAQFEGIVT	Potential
COP9 signalosome complex subunit 6	A6QQ21|CSN6_BOVIN	E100	0.756	0.763	IIDKEYYYT	Potential
COP9 signalosome complex subunit 6	F1MG10|F1MG10_BOVIN	E100	0.756	0.763	IIDKEYYYT	Potential
Peptidyl-glycine alpha-amidating monooxygenase	F1MZB4|F1MZB4_BOVIN	E611	0.756	0.759	PKSKEGPLL	Potential
Peptidyl-glycine alpha-amidating monooxygenase	F1MZN4|F1MZN4_BOVIN	E611	0.756	0.759	PKSKEGPLL	Potential

Peptidyl-glycine alpha-amidating monooxygenase	P10731|AMD_BOVIN	E611	0.756	0.759	PKSKEGPLL	Potential
Transmembrane emp24 domain-containing protein 5	Q2KJ84|TMED5_BOVIN	E77	0.756	0.757	LASPEGKTL	Potential
Sine oculis-binding protein homolog	A7XYH9|SOBP_BOVIN	E7	0.755	0.814	EMEKEGRPP	Potential
Sine oculis-binding protein homolog	G5E5M4|G5E5M4_BOVIN	E7	0.755	0.814	EMEKEGRPP	Potential
MAPK-interacting and spindle-stabilizing protein-like	Q5E9L3|MISSL_BOVIN	E120	0.755	0.76	MPFPELPRP	Potential
Butyrophilin subfamily 1 member A1	E1BHI7|E1BHI7_BOVIN	E197	0.754	0.805	NPDEEGLFT	Potential
Butyrophilin subfamily 1 member A1 (Fragment)	F1MZQ4|F1MZQ4_BOVIN	E199	0.754	0.805	NPDEEGLFT	Potential
Disintegrin and metalloproteinase domain-containing protein 10	Q10741|ADA10_BOVIN	E533	0.754	0.79	DCAKEGICN	Potential
Armadillo repeat-containing protein 2	F1MJ05|F1MJ05_BOVIN	E752	0.754	0.784	LILKEGGGI	Potential
Armadillo repeat-containing protein 2	P0C6R2|ARMC2_BOVIN	E754	0.754	0.784	LILKEGGGI	Potential
Protein FAM110B	E1BF54|E1BF54_BOVIN	E111	0.754	0.772	GAKSEGGAP	Potential
Protein FAM110B	Q2KJ38|F110B_BOVIN	E111	0.754	0.772	GAKSEGGAP	Potential
DNA-(apurinic or apyrimidinic site) lyase 2	F1MSK4|F1MSK4_BOVIN	Q495	0.754	0.77	CARPQGPPT	Potential
DNA-(apurinic or apyrimidinic site) lyase 2	Q5E9N9|APEX2_BOVIN	Q495	0.754	0.77	CARPQGPPT	Potential
Methionine aminopeptidase 2	F1N4Q2|F1N4Q2_BOVIN	E58	0.754	0.765	EPDKEAGAS	Potential
Methionine aminopeptidase 2	Q3ZC89|MAP2_BOVIN	E58	0.754	0.765	EPDKEAGAS	Potential
Histone H2A	Q1LZ92|Q1LZ92_BOVIN	E168	0.754	0.754	DSDKEGTSN	Potential
FGD1 protein	A5D7I5|A5D7I5_BOVIN	E128	0.753	0.831	EPHPEGPQR	Potential
Protein-methionine sulfoxide oxidase MICAL3	G3MWR8|MICA3_BOVIN	E1366	0.753	0.785	LPAAEGAQP	Potential
Nitric oxide synthase	F1MYR5|F1MYR5_BOVIN	E614	0.753	0.771	LMLKELTNT	Potential
Nitric oxide synthase, inducible	Q27995|NOS2_BOVIN	E614	0.753	0.771	LMLKELTNT	Potential
KRT82 protein	A3KMY1|A3KMY1_BOVIN	E168	0.753	0.756	EPIFEGYIS	Potential
LIX1L protein	A2VE91|A2VE91_BOVIN	E311	0.753	0.756	LAGKELRFH	Potential
PAIP1 protein	A5PKG2|A5PKG2_BOVIN	E257	0.753	0.756	DAWKEKGKT	Potential
Ubiquitin carboxyl-terminal hydrolase	E1BKF4|E1BKF4_BOVIN	E968	0.753	0.752	PSPKEGPSA	Potential
HLA-B associated transcript 4	Q0P5E9|Q0P5E9_BOVIN	E274	0.753	0.75	GPRGEGRAN	Potential
Vacuolar protein sorting-associated protein 35	Q2HJG5|VPS35_BOVIN	E787	0.752	0.814	SPESEGPIY	Potential
Oxidative stress induced growth inhibitor 1	Q2KIN7|Q2KIN7_BOVIN	E441	0.752	0.798	TIHQEGLYA	Potential
Neural cell adhesion molecule 1	P31836|NCAM1_BOVIN	E567	0.752	0.783	EASMEGIVT	Potential
Neural cell adhesion molecule 1 (Fragment)	F1MMJ2|F1MMJ2_BOVIN	E567	0.752	0.783	EASMEGIVT	Potential
DNA repair protein RAD51 homolog 1	Q2KJ94|RAD51_BOVIN	E163	0.752	0.764	YIDTEGTFR	Potential
Gamma-aminobutyric acid receptor subunit alpha-4	P20237|GBRA4_BOVIN	E199	0.752	0.76	YPKSEMIYT	Potential
ATF6B protein	A4FUY2|A4FUY2_BOVIN	E290	0.751	0.772	RVQPEGPTP	Potential
Vesicular glutamate transporter 2	A6QLI1|VGLU2_BOVIN	E14	0.751	0.769	TPGKEGLKN	Potential
LOC530205 protein	A4FUB1|A4FUB1_BOVIN	E303	0.751	0.765	HPKPEGLAA	Potential
HERC3 protein	A2VDP7|A2VDP7_BOVIN	E757	0.751	0.764	LLLKELLNP	Potential
SH3RF2 protein	A4IFR5|A4IFR5_BOVIN	E46	0.751	0.763	KAHKELRCP	Potential
Pentatricopeptide repeat-containing protein 2, mitochondrial	Q3SZ55|PTCD2_BOVIN	E89	0.751	0.757	LILKEELRT	Potential

ETS homologous factor	Q32LN0|EHF_BOVIN	E183	0.751	0.754	LPIAESPDT	Potential
Protein NDRG1	F1MS38|F1MS38_BOVIN	E170	0.75	0.812	NPCAEGWMD	Potential
Protein NDRG1	Q3SYX0|NDRG1_BOVIN	E170	0.75	0.812	NPCAEGWMD	Potential
Mitochondrial peptide methionine sulfoxide reductase	P54149|MSRA_BOVIN	E61	0.75	0.805	EPFPEGTQM	Potential
Trifunctional enzyme subunit beta, mitochondrial	O46629|ECHB_BOVIN	E449	0.75	0.802	RLRKEGGQY	Potential
39S ribosomal protein L45, mitochondrial	F1MEP0|F1MEP0_BOVIN	E103	0.75	0.793	SLSKEGLAQ	Potential
39S ribosomal protein L45, mitochondrial	Q3T142|RM45_BOVIN	E103	0.75	0.793	SLSKEGLAQ	Potential
Cathelicidin-1	P22226|CTHL1_BOVIN	E108	0.75	0.786	LKRCEGTVT	Potential
ZW10 interactor	F1N7G9|F1N7G9_BOVIN	E98	0.75	0.781	EQWKELKAT	Potential
ZW10 interactor	Q2TBH8|ZWINT_BOVIN	E98	0.75	0.781	EQWKELKAT	Potential
UPF0488 protein C8orf33 homolog	Q2KID8|CH033_BOVIN	E169	0.75	0.772	RPRPEGRSK	Potential

A.6.2. Predicted porcine targets for FMDV 3C protease

Protein name	Uniprot identifer	P1 amino acid	Cleavage score	Surface score	Predicted cleavage site	Likelihood of cleavage site
Oxysterol-binding protein (Fragment)	I3LKX6|I3LKX6_PIG	E305	0.856	0.764	EPLKETTYT	Potential
Neuroendocrine convertase 2	F1SBI5|F1SBI5_PIG	E64	0.845	0.787	LPFAEGLYH	Potential
Neuroendocrine convertase 2	Q03333|NEC2_PIG	E64	0.845	0.787	LPFAEGLYH	Potential
Aquaporin 6	B2MUK1|B2MUK1_PIG	E263	0.844	0.757	EPQKEGSQA	Potential
Sodium channel protein	F1RSK2|F1RSK2_PIG	E1764	0.842	0.782	APEKEGLIA	Potential
Oxysterol-binding protein	F1RY89|F1RY89_PIG	E60	0.841	0.798	PPRDEGPPT	Potential
Lysoplasmalogenase	F1RMN2|TM86B_PIG	E6	0.839	0.777	DPGKEGLPR	Potential
Complement component MASP3	D5L7X4|D5L7X4_PIG	E580	0.836	0.828	RPEPEGPAP	Potential
Sodium channel protein	F1SJS7|F1SJS7_PIG	E665	0.834	0.768	DPFAELTIT	Potential
Integrin beta-1-binding protein 2	Q462R2|ITBP2_PIG	E76	0.832	0.819	APQPEGPAT	Potential
Galectin	I3LGZ8|I3LGZ8_PIG	E181	0.83	0.777	LPSMEGSPT	Potential
1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4	Q8SPR7|PLCD4_PIG	E657	0.826	0.77	DDSKEGSIV	Potential
Sialoadhesin	A7LCJ3|SN_PIG	E1236	0.825	0.763	RPSDEGLYS	Potential
Dihydropyrimidine dehydrogenase [NADP(+)]	F1S550|F1S550_PIG	E650	0.823	0.767	RAAKEGGAD	Potential
Dihydropyrimidine dehydrogenase [NADP(+)]	Q28943|DPYD_PIG	E726	0.823	0.767	RAAKEGGAD	Potential
V-type proton ATPase subunit F	F1SMN6|F1SMN6_PIG	E95	0.822	0.754	IPSKEHPYD	Potential
Calpain-1 catalytic subunit	P35750|CAN1_PIG	E506	0.819	0.773	EPNKEGDFV	Potential
Aminopeptidase N	F1SK03|F1SK03_PIG	E493	0.818	0.769	DLFKEGLAS	Potential
Aminopeptidase N	P15145|AMPN_PIG	E493	0.818	0.769	DLFKEGLAS	Potential
Heat shock protein beta-1	Q5S1U1|HSPB1_PIG	E65	0.814	0.771	PPAIEGPAA	Potential
Interleukin-12 receptor subunit beta-2	F1S826|F1S826_PIG	E812	0.812	0.76	LPSHEAPIT	Potential
Interleukin-12 receptor subunit beta-2	K7GQN9|K7GQN9_PIG	E726	0.812	0.76	LPSHEAPIT	Potential
Interleukin-12 receptor subunit beta-2	Q8MJS1|I12R2_PIG	E812	0.812	0.76	LPSHEAPIT	Potential
Calpain-7	A0FKG7|CAN7_PIG	E794	0.811	0.79	LPKQEGPFF	Potential
Calpain-7	I3LBF0|I3LBF0_PIG	E794	0.811	0.79	LPKQEGPFF	Potential
Annexin	F1SS97|F1SS97_PIG	E213	0.809	0.806	LKRAEGPST	Potential
Major histocompatibility complex, class II, DM alpha	A5D9K9|A5D9K9_PIG	E117	0.809	0.752	GPLFEGKIP	Potential
Major histocompatibility complex, class II, DM alpha	Q9BEA4|Q9BEA4_PIG	E117	0.809	0.752	GPLFEGKIP	Potential
Nitric oxide synthase, endothelial	Q28969|NOS3_PIG	Q29	0.807	0.771	LCGKQGPAT	Potential
4-trimethylaminobutyraldehyde dehydrogenase	F1S232|F1S232_PIG	E473	0.805	0.771	PKLKEGYYM	Potential
Gastrokine-3	D2XPP7|GKN3_PIG	E150	0.805	0.753	QQQKEGTAL	Potential
Interleukin 16	Q76EI9|Q76EI9_PIG	E457	0.804	0.789	LASQEGTIQ	Potential
WAP four-disulfide core domain protein 2	Q8MI69|WFDC2_PIG	E76	0.803	0.798	PNEKEGSCP	Potential
Mitochondrial uncoupling protein 2	O97562|UCP2_PIG	E264	0.8	0.762	MLQKEGPRA	Potential
Mitogen-activated protein kinase kinase kinase	F1SFP4|F1SFP4_PIG	E646	0.8	0.75	TPPSEGSAP	Potential
Mitogen-activated protein kinase kinase kinase 12	I3L8B0|I3L8B0_PIG	E679	0.8	0.75	TPPSEGSAP	Potential
V-type proton ATPase subunit a	F1RFL9|F1RFL9_PIG	E364	0.799	0.78	IPTKETPPT	Potential
Isoform 3 of Sterol regulatory element-binding protein cleavage-activating protein	Q5MNU5|SCAP_PIG	E948	0.799	0.759	TPEDEGSFP	Potential
Sterol regulatory element-binding protein cleavage-activating protein	F1SNT3|F1SNT3_PIG	E949	0.799	0.759	TPEDEGSFP	Potential
GNAS complex locus	A5GFT7|A5GFT7_PIG	R242	0.796	0.766	SPSKRGPIP	Potential
Endothelin B receptor	P35463|EDNRB_PIG	E92	0.796	0.763	PPPCEGPIE	Potential
Hermansky-Pudlak syndrome 6 protein	A5A775|A5A775_PIG	E400	0.795	0.785	LPSAEGLVF	Potential
Olfactory receptor (Fragment)	I3LFY3|I3LFY3_PIG	E232	0.795	0.775	MPSAEGRAK	Potential
Tyrosine-protein kinase receptor (Fragment)	F1S3X9|F1S3X9_PIG	E621	0.793	0.784	EAFMEGIFT	Potential

Tyrosine-protein kinase receptor (Fragment)	K7GQT6|K7GQT6_PIG	E699	0.793	0.784	EAFMEGIFT	Potential
Protein delta homolog 2	B2LW77|DLK2_PIG	E132	0.793	0.755	CERKEGPCE	Potential
von Willebrand factor	K7GNN0|K7GNN0_PIG	E568	0.793	0.755	NPRAEGLEC	Potential
von Willebrand factor (Fragment)	F1SL22|F1SL22_PIG	E606	0.793	0.755	NPRAEGLEC	Potential
von Willebrand factor (Fragment)	Q28833|VWF_PIG	E458	0.793	0.755	NPRAEGLEC	Potential
Wee1-like protein kinase 2	A4PES0|WEE2_PIG	E109	0.791	0.768	PPHCESPFT	Potential
Tissue factor pathway inhibitor	F1SFC1|F1SFC1_PIG	E42	0.791	0.764	LPPDEGPCR	Potential
Iodotyrosine dehalogenase 1	Q6TA49|IYD1_PIG	E271	0.791	0.758	YPSKEATVP	Potential
Microtubule-associated protein	F1SLI3|F1SLI3_PIG	E479	0.79	0.776	PPNKELPPS	Potential
Chaperonin containing TCP1, subunit 2 (Beta)	D0G0C8|D0G0C8_PIG	E482	0.789	0.782	LDMKEGTIG	Potential
XK-related protein	F1STN1|F1STN1_PIG	E258	0.789	0.764	DPCSEWLYR	Potential
Fibulin-1	F1SM61|F1SM61_PIG	E335	0.789	0.751	CINTEGSYT	Potential
GH3 domain-containing protein	E9KYT5|E9KYT5_PIG	E433	0.788	0.79	LDSSEGSAP	Potential
Xaa-Pro aminopeptidase 2	Q95333|XPP2_PIG	E547	0.787	0.816	IPMAEGMFT	Potential
Protein S100-A12	K7GR24|K7GR24_PIG	E40	0.787	0.791	LITKELPNT	Potential
Protein S100-A12	P80310|S10AC_PIG	E40	0.787	0.791	LITKELPNT	Potential
Oxytocin receptor	E3V2E8|E3V2E8_PIG	E24	0.786	0.78	PPEAEGNRT	Potential
Oxytocin receptor	P32306|OXYR_PIG	E24	0.786	0.78	PPEAEGNRT	Potential
LIM domain only 3	A9ED84|A9ED84_PIG	E138	0.786	0.773	GLMKEGYAP	Potential
Phosphoinositide phospholipase C	F1RQ08|F1RQ08_PIG	E999	0.785	0.775	QARAEGRCR	Potential
Calcitonin gene-related peptide type 1 receptor	Q8WN93|CALRL_PIG	E358	0.785	0.771	PWRPEGKIA	Potential
Phosphoacetylglucosamine mutase	F1RQM2|F1RQM2_PIG	E81	0.785	0.759	DPLGEMLAP	Potential
Prolactin receptor	Q6JTA8|PRLR_PIG	E581	0.785	0.757	EPTKETPPS	Potential
Prolactin receptor (Fragment)	K7GKV2|K7GKV2_PIG	E596	0.785	0.757	EPTKETPPS	Potential
Coronin	F1RK53|F1RK53_PIG	E673	0.783	0.811	EPLQEGPGP	Potential
Autophagy-related protein 13 (Fragment)	F1SIA3|F1SIA3_PIG	E341	0.783	0.757	VPGKEGGVP	Potential
Coatomer subunit gamma	F1SNE9|F1SNE9_PIG	E462	0.782	0.821	LLGKEGPRT	Potential
Apolipoprotein M	Q2LE37|APOM_PIG	E58	0.782	0.765	APTKEELAT	Potential
Toll-like receptor 8	B3XXC2|B3XXC2_PIG	E1000	0.781	0.796	NPKAEGLFW	Potential
Pulmonary surfactant-associated protein D	Q9N1X4|SFTPD_PIG	E120	0.781	0.77	PAGKEGPSG	Potential
Bifunctional ATP-dependent dihydroxyacetone kinase/FAD-AMP lyase (cyclizing)	F1RKQ4|DHAK_PIG	E128	0.78	0.775	QARAEGIPV	Potential
Inhibin beta A chain	F1SSD8|F1SSD8_PIG	E143	0.78	0.773	EISKEGSDL	Potential
Inhibin beta A chain	P03970|INHBA_PIG	E143	0.78	0.773	EISKEGSDL	Potential
Iron-responsive element-binding protein 2	B3VKQ2|IREB2_PIG	E838	0.78	0.75	LYQKEGIPL	Potential
Prenylated Rab acceptor protein 1	F1RMX2|F1RMX2_PIG	E14	0.779	0.79	DAEAEGLSA	Potential
Prenylated Rab acceptor protein 1	Q52NJ0|PRAF1_PIG	E14	0.779	0.79	DAEAEGLSA	Potential
Histidine ammonia-lyase	F1SQR7|F1SQR7_PIG	E301	0.779	0.764	LKPKEGLAL	Potential
Glutathione S-transferase omega-1	Q9N1F5|GSTO1_PIG	E21	0.777	0.753	GPVPEGLIR	Potential
Coiled-coil and C2 domain containing 1A	F1SD63|F1SD63_PIG	E934	0.776	0.768	EAAKEALYR	Potential
Histone acetyltransferase	F1RIR8|F1RIR8_PIG	E32	0.776	0.752	NAAVEGTAP	Potential
Histone acetyltransferase	K7GPZ7|K7GPZ7_PIG	E32	0.776	0.752	NAAVEGTAP	Potential
Eosinophil peroxidase	K7GNQ2|K7GNQ2_PIG	E21	0.774	0.794	AWSCEGTAP	Potential
Eosinophil peroxidase (Fragment)	F1RSB4|F1RSB4_PIG	E26	0.774	0.794	AWSCEGTAP	Potential
Phosphoinositide phospholipase C (Fragment)	I3LCQ7|I3LCQ7_PIG	E295	0.774	0.753	EPSKEGQEK	Potential
Protein arginine N-methyltransferase 5	C3RZ98|C3RZ98_PIG	E499	0.771	0.776	EAQFEMPYV	Potential
POU domain protein	F1RM14|F1RM14_PIG	E58	0.771	0.764	LARAEALAA	Potential
POU domain protein	I3L5L7|I3L5L7_PIG	E61	0.771	0.764	LARAEALAA	Potential
POU domain protein	I3LTT4|I3LTT4_PIG	E59	0.771	0.764	LARAEALAA	Potential
POU domain protein (Fragment)	F1RS67|F1RS67_PIG	E18	0.771	0.764	LARAEALAA	Potential
Olfactory receptor	I3LKS3|I3LKS3_PIG	E232	0.771	0.757	MKSKEGRIK	Potential
Olfactory receptor	F1RMQ7|F1RMQ7_PIG	E100	0.771	0.752	LISFEGCMT	Potential
Myosin-1	F1SS64|F1SS64_PIG	E900	0.77	0.788	QAEAEGLAD	Potential

Myosin-2	Q9TV63|MYH2_PIG	E900	0.77	0.788	QAEAEGLAD	Potential
Cytochrome c oxidase subunit 1	O79876|COX1_PIG	E40	0.77	0.759	LIRAELGQP	Potential
Acyl-CoA desaturase	K7GKT7|K7GKT7_PIG	E63	0.769	0.814	YQDKEGPRP	Potential
Acyl-CoA desaturase	Q6RWA7|Q6RWA7_PIG	E63	0.769	0.814	YQDKEGPRP	Potential
Acyl-CoA desaturase (Fragment)	K7GNB9|K7GNB9_PIG	E25	0.769	0.814	YQDKEGPRP	Potential
Adenosine monophosphate deaminase 1 isoform M	B5SYT7|B5SYT7_PIG	E227	0.767	0.76	AASKEEPKP	Potential
Olfactory receptor	I3LIR8|I3LIR8_PIG	E230	0.766	0.798	IPSAEGRAK	Potential
Olfactory receptor	I3LL29|I3LL29_PIG	E230	0.766	0.798	IPSAEGRAK	Potential
Serine/threonine-protein kinase PLK	F1S348|F1S348_PIG	E389	0.766	0.787	DARPEAPAA	Potential
Polymeric immunoglobulin receptor	F1SEY8|F1SEY8_PIG	E71	0.766	0.757	LISSEGYIS	Potential
Insulin-like 3	F1SMC2|F1SMC2_PIG	E213	0.764	0.797	LIAQEGPFF	Potential
Transporter	I3L6B0|I3L6B0_PIG	E570	0.763	0.767	LCQTEGPFL	Potential
Transporter (Fragment)	F1SPI3|F1SPI3_PIG	E573	0.763	0.767	LCQTEGPFL	Potential
Citrate synthase, mitochondrial	P00889|CISY_PIG	E117	0.762	0.828	EPLPEGLFW	Potential
Diacylglycerol kinase (Fragment)	I3LQV4|I3LQV4_PIG	E814	0.762	0.794	DVRAEGTPA	Potential
Caspase	I3L9A3|I3L9A3_PIG	E126	0.762	0.793	EPRPEAPRP	Potential
Taste receptor type 2	I3LB59|I3LB59_PIG	E170	0.762	0.771	NPFKEMPPL	Potential
Ribonucleoside-diphosphate reductase	I3LEK6|I3LEK6_PIG	E60	0.762	0.762	DDSIEGIYD	Potential
2-amino-3-carboxymuconate-6-semialdehyde decarboxylase (Fragment)	F1S0C2|F1S0C2_PIG	E101	0.759	0.754	RCVKELGFP	Potential
Asparagine synthetase	D0G0C6|D0G0C6_PIG	E514	0.758	0.813	PKTKEGYYY	Potential
Isoform 2 of Large proline-rich protein BAG6	A5D9M6|BAG6_PIG	E989	0.758	0.774	PPAPEGGSR	Potential
Large proline-rich protein BAG6	A5D9M6-2|BAG6_PIG	E988	0.758	0.774	PPAPEGGSR	Potential
Large proline-rich protein BAG6	F1RQX8|F1RQX8_PIG	E989	0.758	0.774	PPAPEGGSR	Potential
Liver X receptor alpha	Q4TU03|Q4TU03_PIG	K86	0.757	0.806	QKRKKGPAP	Potential
Syntaxin-binding protein 2	F1SCI9|F1SCI9_PIG	E53	0.757	0.771	DILAEGITI	Potential
Syntaxin-binding protein 2 (Fragment)	Q29268|STXB2_PIG	E53	0.757	0.771	DILAEGITI	Potential
Stress-70 protein, mitochondrial	F1RGJ3|F1RGJ3_PIG	E352	0.756	0.771	RAQFEGIVT	Potential
COP9 signalosome complex subunit 6	A7TX81|CSN6_PIG	E99	0.756	0.763	IIDKEYYYT	Potential
Ubiquitin-conjugating enzyme E2L 6	F4ZS20|F4ZS20_PIG	E17	0.755	0.758	DLQKELPRY	Potential
One cut domain family member	F1SDH9|F1SDH9_PIG	E394	0.755	0.755	RPSKEMQVT	Potential
Histone H2A (Fragment)	I3LQZ2|I3LQZ2_PIG	E146	0.754	0.754	DSDKEGTSN	Potential
DNA repair protein RAD51	B0M1M6|B0M1M6_PIG	E163	0.752	0.764	YIDTEGTFR	Potential
Galectin-4	Q29058|LEG4_PIG	E183	0.75	0.819	LPCMEGAPT	Potential

A.7. ToppFun analysis

Results from the ToppFun analysis of potential FMDV 3C targets in the bovine and porcine proteomes are presented here. As described in Chapter 5, no information for either Biological Process or Molecular Function could be given for the bovine data. Therefore, only the Pathway data is presented.

A.7.1. Bovine targets - pathway

ID	Name	p-value	q-value Bonferroni	q-value FDR B&H	q-value FDR B&Y	Hit Count in Query List	Hit Count in Genome
82927	Citrate cycle (TCA cycle)	2.80E-05	1.75E-02	1.47E-02	1.03E-01	4	30
835393	conversion of glucose to acetyl CoA and entry into the TCA cycle	6.54E-05	4.09E-02	1.47E-02	1.03E-01	4	37
790012	Biosynthesis of amino acids	7.05E-05	4.41E-02	1.47E-02	1.03E-01	5	73
MAP00020_Citrate_cycle_TCA_cycle	MAP00020 Citrate cycle TCA cycle	1.55E-04	9.68E-02	1.94E-02	1.36E-01	3	18
SMP00057	Citric Acid Cycle	1.55E-04	9.68E-02	1.94E-02	1.36E-01	3	18
PW:0000026	citrate cycle	2.14E-04	1.34E-01	1.99E-02	1.40E-01	3	20
814926	Carbon metabolism	2.23E-04	1.39E-01	1.99E-02	1.40E-01	5	93
PW:0000015	Alzheimer disease	3.74E-04	2.34E-01	2.60E-02	1.83E-01	3	24
83002	Glyoxylate and dicarboxylate metabolism	3.74E-04	2.34E-01	2.60E-02	1.83E-01	3	24
198820	Acetylcholine Synthesis	7.21E-04	4.51E-01	4.10E-02	2.88E-01	2	7
P02772	Pyruvate metabolism	7.21E-04	4.51E-01	4.10E-02	2.88E-01	2	7
413347	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	9.57E-04	5.99E-01	4.61E-02	3.23E-01	2	8
M6831	The Citric Acid Cycle	9.57E-04	5.99E-01	4.61E-02	3.23E-01	2	8

A.7.2. Porcine targets - biological process

ID	Name	p-value	q-value Bonferroni	q-value FDR B&H	q-value FDR B&Y	Hit Count in Query List	Hit Count in Genome
GO:0040013	negative regulation of locomotion	2.52E-05	3.61E-02	2.71E-02	2.12E-01	5	212
GO:0032966	negative regulation of collagen biosynthetic process	5.69E-05	8.13E-02	2.71E-02	2.12E-01	2	7
GO:0006101	citrate metabolic process	5.69E-05	8.13E-02	2.71E-02	2.12E-01	2	7
GO:0010713	negative regulation of collagen metabolic process	7.58E-05	1.08E-01	2.71E-02	2.12E-01	2	8
GO:0044252	negative regulation of multicellular organismal metabolic process	9.73E-05	1.39E-01	2.78E-02	2.18E-01	2	9
GO:2000674	regulation of type B pancreatic cell apoptotic process	1.22E-04	1.74E-01	2.90E-02	2.27E-01	2	10
GO:0042640	anagen	1.78E-04	2.54E-01	3.18E-02	2.49E-01	2	12
GO:0097050	type B pancreatic cell apoptotic process	1.78E-04	2.54E-01	3.18E-02	2.49E-01	2	12
GO:0044093	positive regulation of molecular function	2.03E-04	2.90E-01	3.23E-02	2.53E-01	10	1616
GO:0008284	positive regulation of cell proliferation	3.31E-04	4.73E-01	4.18E-02	3.28E-01	7	816
GO:0072350	tricarboxylic acid metabolic process	3.65E-04	5.21E-01	4.18E-02	3.28E-01	2	17
GO:0021537	telencephalon development	3.79E-04	5.43E-01	4.18E-02	3.28E-01	4	207
GO:0030900	forebrain development	3.91E-04	5.60E-01	4.18E-02	3.28E-01	5	380
GO:0044851	hair cycle phase	4.10E-04	5.86E-01	4.18E-02	3.28E-01	2	18
GO:0046597	negative regulation of viral entry into host cell	4.57E-04	6.54E-01	4.27E-02	3.35E-01	2	19
GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity	5.03E-04	7.19E-01	4.27E-02	3.35E-01	4	223
GO:0048820	hair follicle maturation	5.08E-04	7.26E-01	4.27E-02	3.35E-01	2	20
GO:0044711	single-organism biosynthetic process	5.57E-04	7.96E-01	4.42E-02	3.47E-01	9	1491
GO:1902186	regulation of viral release from host cell	6.16E-04	8.81E-01	4.64E-02	3.64E-01	2	22

A.7.3. Porcine targets - molecular function

ID	Name	p-value	q-value Bonferroni	q-value FDR B&H	q-value FDR B&Y	Hit Count in Query List	Hit Count in Genome
GO:0046965	retinoid X receptor binding	4.96E-04	1.30E-01	4.18E-02	2.57E-01	2	19
GO:0004108	citrate (Si)-synthase activity	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0070379	high mobility group box 1 binding	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0070539	linoleic acid binding	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0050785	advanced glycation end-product receptor activity	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0070551	endoribonuclease activity, cleaving siRNA-paired mRNA	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0016515	interleukin-13 receptor activity	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0030233	deoxynucleotide transmembrane transporter activity	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0016501	prostacyclin receptor activity	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0036440	citrate synthase activity	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1
GO:0045503	dynein light chain binding	1.75E-03	4.60E-01	4.18E-02	2.57E-01	1	1

A.7.4. Porcine targets - pathway

ID	Name	p-value	q-value Bonferroni	q-value FDR B&H	q-value FDR B&Y	Hit Count in Query List	Hit Count in Genome
83002	Glyoxylate and dicarboxylate metabolism	1.54E-05	4.55E-03	4.55E-03	2.85E-02	3	24
413347	Citrate cycle, first carbon oxidation, oxaloacetate => 2-oxoglutarate	1.15E-04	3.39E-02	1.13E-02	7.09E-02	2	8
M6831	The Citric Acid Cycle	1.15E-04	3.39E-02	1.13E-02	7.09E-02	2	8
MAP00630_Glyoxy	MAP00630 Glyoxylate and dicarboxylate metabolism	1.84E-04	5.44E-02	1.36E-02	8.52E-02	2	10
790012	Biosynthesis of amino acids	4.43E-04	1.31E-01	1.83E-02	1.15E-01	3	73
782397	TCA cycle	4.88E-04	1.44E-01	1.83E-02	1.15E-01	2	16
198787	TCA Cycle	5.52E-04	1.63E-01	1.83E-02	1.15E-01	2	17
714485	2-Oxocarboxylic acid metabolism	5.52E-04	1.63E-01	1.83E-02	1.15E-01	2	17
MAP00020_Citrat	MAP00020 Citrate cycle TCA cycle	6.20E-04	1.83E-01	1.83E-02	1.15E-01	2	18
SMP00057	Citric Acid Cycle	6.20E-04	1.83E-01	1.83E-02	1.15E-01	2	18
PW:0000026	citrate cycle	7.69E-04	2.27E-01	2.06E-02	1.29E-01	2	20
105919	Citric acid cycle (TCA cycle)	8.48E-04	2.50E-01	2.09E-02	1.31E-01	2	21
855811	Citrate cycle (TCA cycle, Krebs cycle)	9.32E-04	2.75E-01	2.12E-02	1.33E-01	2	22
82927	Citrate cycle (TCA cycle)	1.74E-03	5.13E-01	3.62E-02	2.27E-01	2	30
PW:0000250	peptidoglycan biosynthetic	2.09E-03	6.15E-01	3.62E-02	2.27E-01	1	1
PW:0000033	energy metabolic	2.09E-03	6.15E-01	3.62E-02	2.27E-01	1	1
142149	glutamine biosynthesis I	2.09E-03	6.15E-01	3.62E-02	2.27E-01	1	1
835393	conversion of glucose to acetyl CoA and entry into the TCA cycle	2.64E-03	7.78E-01	4.32E-02	2.71E-01	2	37

A.8. Publication - Genome editing in sheep and cattle

Transgenic Res
DOI 10.1007/s11248-014-9832-x

ORIGINAL PAPER

Genome edited sheep and cattle

Chris Proudfoot · Daniel F. Carlson · Rachel Huddart · Charles R. Long · Jane H. Pryor · Tim J. King · Simon G. Lillico · Alan J. Mileham · David G. McLaren · C. Bruce A. Whitelaw · Scott C. Fahrenkrug

Received: 11 July 2014 / Accepted: 26 August 2014

Abstract Genome editing tools enable efficient and accurate genome manipulation. An enhanced ability to modify the genomes of livestock species could be utilized to improve disease resistance, productivity or breeding capability as well as the generation of new biomedical models. To date, with respect to the direct injection of genome editor mRNA into livestock zygotes, this technology has been limited to the generation of pigs with edited genomes. To capture the far-reaching applications of gene-editing, from disease modelling to agricultural improvement, the technology must be easily applied to a number of species using a variety of approaches. In this study, we demonstrate zygote injection of TALEN mRNA can also produce gene-edited cattle and sheep. In both species we have targeted the myostatin (*MSTN*) gene. In addition, we report a critical innovation for application of gene-editing to the cattle industry whereby gene-edited calves can be produced with specified genetics by ovum pickup, in vitro fertilization and zygote microinjection (OPU-IVF-ZM). This provides a practical alternative to somatic cell nuclear transfer for gene knockout or introgression of desirable alleles into a target breed/genetic line.

Keywords Livestock · TALEN · Myostatin · Zygote · Genetic engineering

Chris Proudfoot and Daniel F. Carlson are joint first authors.

C. Proudfoot · R. Huddart · T. J. King · S. G. Lillico · C. B. A. Whitelaw (✉)
The Roslin Institute and R(D)SVS, University of Edinburgh, Easter Bush Campus, Edinburgh EH25 9RG, UK
e-mail: bruce.whitelaw@roslin.ed.ac.uk

D. F. Carlson · S. C. Fahrenkrug (✉)
Recombinetics Inc, 2575 University Ave. West, Suite 100, Saint Paul, MN 55108, USA
e-mail: scott@recombinetics.com

C. R. Long · J. H. Pryor
Department of Veterinary Physiology and Pharmacology, College of Veterinary Medicine, Texas A&M University, College Station, TX 77843, USA

A. J. Mileham · D. G. McLaren
Genus plc, 1525 River Road, DeForest, WI 53532, USA

Introduction

The ability to generate gene knockouts is an extremely powerful tool for the analysis of gene function and for the generation of animals with biotechnological or breeding applications (Fahrenkrug et al. 2010). In livestock species this process traditionally involves the generation of a knock-out cell line generated utilising homologous recombination followed by somatic cell nuclear transfer (SCNT). This remains the method of choice for many applications (Kurome

Published online: 10 September 2014

et al. 2013), however, application of SCNT strategies requires a high-level of technical expertise, reliable supply of oocytes and a large recipient herd, features not available in many areas where gene-editing might have the greatest impact.

The advent of highly efficient genome editors has driven a renaissance in livestock genetic modification by embryo microinjection (Tan et al. 2012; Lillico et al. 2013). Whereas pronuclear injection, the original method for creation of transgenic livestock, is rarely performed nowadays due to the low frequency of transgenic offspring (Clark and Whitelaw 2003; Clark et al. 2007; Ivics et al. 2014) cloning strategies are still widely utilised and in combination with zinc finger nucleases have been used to generate edited cattle (Liu et al. 2014), pigs (Hauschild et al. 2011), sheep (Zhang et al. 2014) and goats (Boulanger et al. 2014). In comparison to cloning, cytoplasmic injection of zygotes with editor mRNA is both technically simple and efficient (Geurts et al. 2009; Carbery et al. 2010; Carlson et al. 2012). In this study, we build on our recent success of gene-editing in pigs (Lillico et al. 2013) to derive the first genome edited sheep and cattle. As with our swine study, the editing events were the result of direct injection of TALEN mRNA into zygotes followed by transfer into synchronized recipients. Another critical innovation in this study is that bovine embryos were prepared by ovum pickup, in vitro fertilisation and zygote microinjection (OPU-IVF-ZM). OPU-IVF is widely used in the cattle industry to rapidly produce a number of offspring from a single cow of elite genetics, up to 50–100 offspring per year (Leeuw 2006). Thus, in vitro produced embryos from either in vitro or in vivo matured oocytes can be used for rapid introgression of gene-edits into defined populations.

Materials and methods

TALEN design and construction

Design and construction of the btGDF83.1L+83.1NR is described in Carlson et al. 2012 using the RCIscript-GoldyTALEN transcription vector (Addgene ID 38142). Messenger RNA was synthesized using the mMessage Machine T3 Kit (Ambion) as previously described (Carlson et al. 2012) prior to polyadenylation using the Poly(A) tailing kit (Ambion) according to the manufacturers recommendations. To test the activity of the TALEN pair, 1 μg of mRNA was transfected (Neon^{TM}, Life Technologies; 1800 V, 20 ms, 1 pulse) into 10^6 primary bovine and ovine fibroblasts. The cells were allowed to recover at 30 °C for 3 days before being harvested and the extent of genome modification assessed by Surveyor nuclease assay. The primer pair *MSTN* For (5′-GTCAAGGTAACAGACACACC-3′) and *MSTN* Rev (5′-CACCCACAGCGATCTACTAC-3′) was used to amplify a 359 base pair region both the bovine and ovine Surveyor assays.

Bovine oocyte collection and manipulation

Oocytes were collected from Nellore cows under ultrasound guidance (Aloka 500 and a vaginal guide probe) with a 17 gauge needle connected via a Cook pump set at 72 psi. Oocytes were rinsed with pre-warmed TL Hepes (0.3 % BSA) + Gentamicin (50 μg/μl) supplemented with 10 IU/ml of Heparin and placed into maturation medium. In vitro fertilization was conducted in pre-equilibrated modified Tyrode-lactate medium supplemented with 250 μM sodium pyruvate, 1 % P/S, 6 mg/ml BSA Fatty Acid free (Sigma), 20 μM Penicillamine, 10 μM Hypotaurine, 1 μM Epinephrine and 10 μg/ml Heparin (Sigma) at 38.5 °C, 5 % CO_2 in an air humidified incubator. Frozen semen from a Nelore bull was thawed at 35 °C then separated by centrifugation at 200×*g* in a density gradient medium (Isolate®, Irvine Scientific, Santa Ana, CA, USA) 50 % upper/90 % lower. The sperm pellet was re-suspended in 2 ml of modified Tyrodes medium and centrifuged at 200×*g* for 10 min to wash. This was repeated once more before the sperm pellet was removed and placed into a warm 0.65 ml microtube (VWR Scientific, Pittsburgh, PA., USA). Fertilization was conducted in a Nunclon® 4-well multi-dish (VWR) containing up to 50 matured oocytes per well and a concentration of 1.5 × 10^6 sperm/ml + 20 μg/ml heparin. Presumptive zygotes, 20–22 h post fertilization, were vortexed and further cleaned with a stripper pipette (125 μm diameter) prior to placing in Hanks 199 + 10 % FBS + Gentamicin for an injection of either 2 or 5 ng/μl TALEN mRNA. Injections were conducted under positive pressure until a slight expansion of the cell membrane was observed. All injected zygotes were placed in Evolve + 4 mg/ml Probumin (BSA) + Gentamicin in 5/5/90 humidified

incubator at 38.5 °C. On day 2 post IVF, all non-cleaved embryos were removed leaving the remainder to culture undisturbed until day 7. At 7 days post IVF viable embryos were washed through Vigro Holding Medium (Bioniche), loaded and transferred into synchronized cross-bred recipients.

Ovine oocyte manipulation and transfer

Ovine ovaries were collected from the abattoir and the follicles aspirated with pre-warmed phosphate buffered saline at 38 °C. Oocytes were washed three times in oocyte wash medium before transfer to maturation medium for 22 h (38.5 °C, 5 % CO_2). The oocytes are then washed twice in fertilisation medium before being transferred to a Nunclon® 4-well multidish (VWR) with each well containing 450 ul of fertilisation medium and approximately 40 oocytes. 1×10^6 sperm was added to each well and incubated for 24 h (38.5 °C, 5 % CO_2). The fertilized oocytes were then washed in SOFaaBSA to remove sperm and any remaining cumulus cells. The zygotes were subjected to a single 2–5pl injection of 2 ng/ul TALEN mRNA before being returned to 4 well plates containing 800 ul SOFaaBSA medium per well and cultured in groups of approximately 40 zygotes. The zygotes were incubated (5 % CO_2, 5 % O_2, 90 % N_2, 38.5 °C) for 6–7 days at which point blastocysts were transferred to recipient ewes. Progesterone sponges (Chronogest sponges) were inserted into ewes for a period of between 11 and 15 days. After removal of the sponges the ewes showed estrus 36–48 h later. Schedules were arranged such that day 6 blastocysts were transferred to recipient ewes 6 days post estrus under general anaesthesia, following a mid-line laparotomy to expose the uterus. A Drummond pipette was used to transfer two or three blastocysts into the uterine horn.

Genotyping

Gene editing events were characterized by PCR, Surveyor assays and sequencing as described previously (Carlson et al. 2012). Analysis of bovine samples utilized the primer pair btGDF8 forward (5′-CCTTGAGGTAGGAGAGTGTTTTGGG-3′) and btGDF8 reverse (5′-CTCATGAACACCCACAGCGATCTAC-3′). The lambs were analysed using the primer pair *MSTN* For (5′-GTCAAGGTAACAGACACACC-3′) and *MSTN* Rev (5′-CACCCACAGCGATCTACTAC-3′).

Results and discussion

The aim of this paper was to determine the potential of genome editors as a tool for introducing desired mutations in sheep and cattle species. The *MSTN* gene (McPherron et al. 1997) was considered an attractive target as mutations have been found naturally in both cattle (Grobet et al. 1997) and sheep (Clop et al. 2006). *MSTN* or growth and differentiation factor 8 (GDF-8) is a member of the transforming growth factor β family and is a negative regulator of muscle growth. An 11 bp deletion in the bovine *MSTN* gene at position 821 has been shown to result in muscular hypertrophy or the ‘double muscle’ phenotype characterised by a 20 % increase in muscle mass (Grobet et al. 1997). Mutations in the *MSTN* gene that result in either inactivation or reduction of functional protein also result in a marked increase in muscle mass (McPherron and Lee 1997). Indel formation induced by TALEN activity would be ideal for attempting to replicate the double muscle phenotype and proving the functionality or editors in sheep and cattle. In this particular example, we hypothesized that the easy calving of Nelore cattle will reduce the management burden of dystocia that is common in other cattle breeds with double muscling breeds (Ménissier 1976). For sheep, higher survivability of offspring derived from a terminal Texel (*MSTN*-KO) sire (Leymaster and Jenkins 1993) would be more beneficial in alternative breeds of sheep. Gene editing represents a rapid methodology to introgress *MSTN* inactivating alleles into naive breeds.

We have previously demonstrated activity of TALENs that targeted a region of the bovine *MSTN* gene between positions 815 and 872 relative to the start codon (NM_001001525; Fig. 1a) (Carlson et al. 2012). The left TALEN recognises 23 base pairs and the right TALEN recognises 19 base pairs. Comparison of the bovine and ovine (NM_001009428) sequences showed there to be a SNP at base 8 of the binding site of the left TALEN monomer (Fig. 1a). We have observed that at least one mismatch between the TALENs and the target sequence can be tolerated (Tan et al. 2013), so we hypothesized that the bovine TALENs would also have activity in ovine cells.

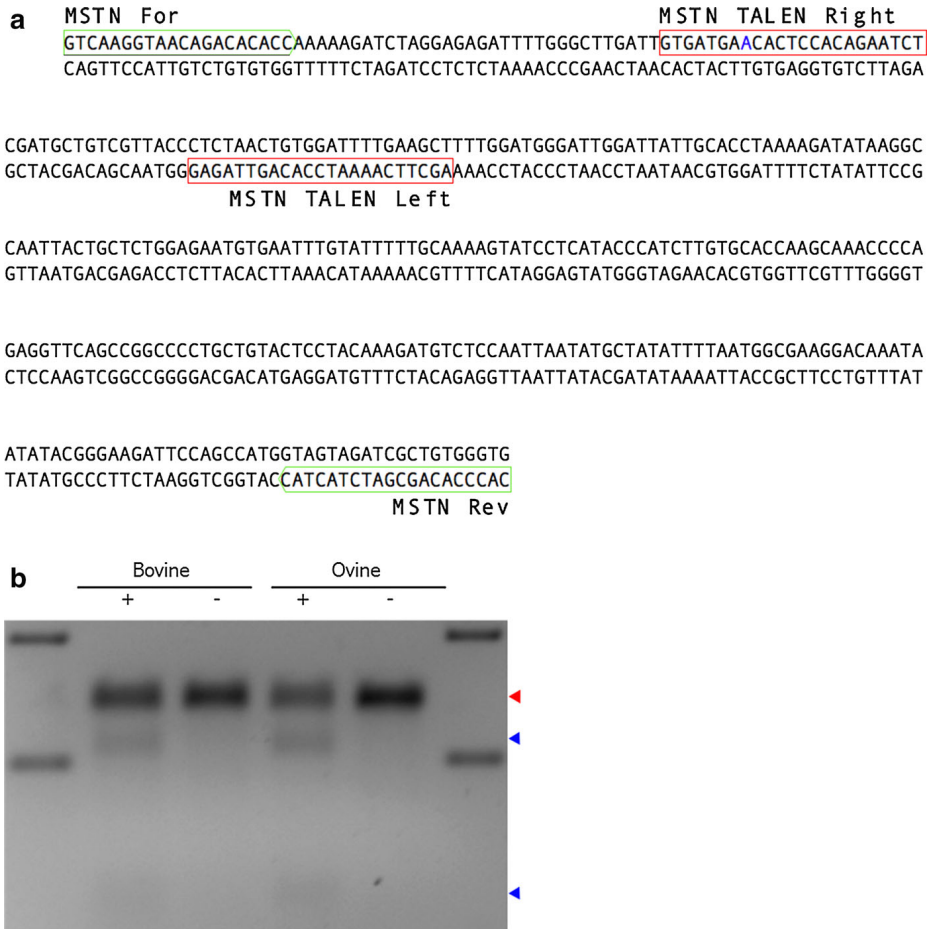


Fig. 1 The *MSTN* TALENs. **a** 359 bp of the bovine *MSTN* gene sequence showing the TALEN binding sites (*red boxes*) and the primers (*green boxes*) used to amplify the region for the surveyor nuclease assay. The base coloured *blue* dictates the position of the mismatch in the ovine sequence in which it is G rather than A. **b** The surveyor nuclease assay results for the TALEN transfected bovine and ovine fibroblasts. gDNA extracted from transfected cells was treated with and without nuclease. (Color figure online)

Indeed, transient transfection of TALEN mRNA into bovine and ovine fibroblasts and subsequent Surveyor nuclease assay showed similar levels of activity in both species (Fig. 1b).

Bovine zygote injections and transfers

Two rounds of OPU-IVF were conducted using Nelore donors in parallel to the same procedure with abattoir oocytes (TransOva Genetics, Sioux Center, IA). After IVF, presumptive zygotes were injected with 2 or 5 ng/µl of TALEN mRNA. Embryos were scored for blastocyst formation on day 7 and a portion of embryos were analysed for gene-edits to evaluate performance of the approach. For embryos derived from abattoir oocytes, 4/13 (31 %) and 17/30 (57 %) were edited for the 2 and 5 ng/µl injections, respectively. Four of the six Nelore blastocysts included in this analysis were edited. In total, 20 Nelore embryos were transferred to 11 outbred recipients resulting in two full term twin-pregnancies (Table 1). The first set of twins birthed naturally resulting in a bull (bull #1) and heifer calf (Fig. 2a). Unfortunately, the second recipient went into labour shortly after a routine check and the birth was unattended. Both bull calves were born dead due to calving difficulties associated with twinning. Sequence analysis revealed that each of the three bull calves was edited whereas no edits were observed in the heifer. A total of 3 unique alleles were sequenced from bull # 1. The predominant genotype 844del1 (70 % of sequenced alleles, n = 13) is a frame shift mutation that results in a stop codon within four amino acids. A second mutant allele, ΔR283, was also observed twice as was the wild type sequence (Fig. 3). This suggests that the TALENs were active for more than one cell division, an observation made previously by analysis of pre-implantation embryos (Carlson et al. 2012). The ΔR283 mutation has not been characterized previously, and since it is not a frame-shift it is expected to produce a protein, but with

Table 1 The development and editing frequency of bovine and ovine zygotes

Dose	Species	Oocytes	Cleaved (%)	Blastocysts (%)	Transferred	Preg (%)	Edited (%)
C	Transova	83	62 (74)	27 (33)	–		
TE	Transova	119	89 (75)	24 (20)	–		
2 ng/μl	Transova	45	34 (76)	12 (27)	–		
	Nelore	21	17 (81)	6 (29)	4	0/2	–
	Sheep	113	61 (54)	27 (24)	26	8/9 (89)	1/9 (11)
5 ng/μl	Transova	308	234 (76)	44 (14)	–		
	Nelore	166	112 (67)	13 (8)	16[a]	2/9 (22)	3/4 (75)

A comparison of the zygote data, pregnancy rates and editing frequencies

C uninjected controls, *TE* TE injected controls

[a] 9 of the transferred were morulae

Fig. 2 *MSTN* edited animals. **a** The live born bull (bull #1: *left*) and heifer calf (*right*). **b** The readily observed phenotypic difference between bull #1 (*right*) and the wild-type heifer (*left*). **c** The edited lamb

Fig. 3 The *MSTN* editing events. An alignment of the bovine and ovine WT sequences and the alleles present in each of the edited animals. The TALEN binding sites are highlighted on the WT sequences, the ovine mismatch is *underlined* and the corresponding amino acid change is indicated on the *right*

```
Nelore WT          GTGATGAACACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGA

Bull 1 Allele 1    GTGATGAACACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGA   WT
Bull 1 Allele 2    GTGATGAACACTCCACAGAATCTCGATGCTGT---TACCCTCTAACTGTGGATTTTGA   ΔR283
Bull 1 Allele 3    GTGATGAACACTCCACAGAATCTCGATGC-GTCGTTACCCTCTAACTGTGGATTTTGA   Δ1

Heifer Allele 1    GTGATGAACACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGA   WT
Heifer Allele 2    GTGATGAACACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGA   WT

Bull 2 Allele 1    GTGATGAACACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGA   WT
Bull 2 Allele 2    GTGATGAACACTCCACAGAATCTCGA---TGTCGTTACCCTCTAACTGTGGATTTTGA   ΔC281

Bull 3 Allele 1    GTGATGAACACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGA   WT
Bull 3 Allele 2    GTGATGAACACTCCACAGAATCTCGA----------------------AGGACAG---   Δ219 +7


Sheep WT           GTGATGAGCACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGA

Sheep Allele 1     GTGATGAGCACTCCACAGAATCTCGATGCTGTCGTTACCCTCTAACTGTGGATTTTGA   WT
Sheep Allele 2     GTGATGAGCACTCCACAGAATCTCGATGCTGT---TACCCTCTAACTGTGGATTTTGA   ΔR283
```

unknown functionality. Regardless, a phenotypic difference between bull #1 and the wild-type heifer is readily observed (Fig. 2b). Given that this bull is mosaic, without segregation it is difficult to differentiate whether hypermuscularity in the Nelore bull derives from haploinsufficiency, from homozygosity of 844del1 knockout allele, or heterozygosity 844del1 and ΔR283 alleles. The potential of ΔR283 as a hypomorphic allele will be evaluated in subsequent generations due to the desire to identify myostatin

genotypes that balance enhanced muscle hypertrophy with calving ease (Keele and Fahrenkrug 2001). Future analysis will also measure the effect of the mosaicism on germline transmission of the editing events from the Nelore bull. Given this, a full analysis of the degree of mosaicism in different tissues, to assess whether or not it is lower in muscle, has been ruled out at this stage.

Ovine zygote injections and transfers

Ovine oocytes were collected from abattoir-derived material and subjected to in vitro maturation and in vitro fertilisation (Ritchie et al. 2008) before receiving a single 2–5pl injection of TALEN mRNA at 2 ng/µl. The zygotes were subsequently cultured for a further 6–7 days before transfer of blastocysts to synchronised recipient ewes. The sheep zygotes showed a good blastocyst development rate of 24 %, despite an initially poorer than expected cleavage rate (Table 1). In total 26 blastocysts were transferred to 9 recipient ewes (2 or 3 blastocysts per ewe) resulting in 8 pregnancies and 12 live births. Three of these lambs died within 24 h post-partum and carcasses were disposed of before samples could be acquired for analysis.

Of the 9 live births one was shown to be edited (Fig. 2c) as a heterozygote ΔR283 (Fig. 3), demonstrating cross-species application and surprisingly an identical genotypic outcome of TALENs designed against the bovine myostatin gene. As with bull #1, the sequence context does not enable definitive identification as to which three bases have been deleted from the MSTN gene. For example, in Fig. 3, the three deleted bases have been marked as those coding for R283, alternatively, the three-base deletion could equally have started 1 or 2 bases downstream. However, in all three scenarios the resulting nucleotide and amino acid sequences would be the same.

Conclusion

This study further exemplifies the utility and ease with which TALENs can be used to engineer the genome of livestock. Specifically we demonstrate that sheep and cattle can be added to the growing list of species for which genome editing is now practical. It is anticipated that these tools will accelerate the utilisation of engineered livestock for biomedical and agricultural applications. Genome edited livestock differ from traditional GM animals in that no recombinant DNA (transgene) is integrated into the animal genome. Combined with the ability to mimic desirable or pre-existing mutations, genome editing overcomes many of the issues associated GM animals increasing the likelihood for societal acceptance. Furthermore, the advent of this technology is extremely timely given the global challenge of food security; targeted mutagenesis and allele introgression has the potential to accelerate genetic advancement of agriculturally important traits. The deployment of gene editing using industry-standard reproductive technologies demonstrates several practical approaches to advancing livestock genetics and biotechnology.

Acknowledgments CP, SGL and CBAW receive support from BBSRC ISPG and Grant BB/L007371/1; and from Genus plc. RH holds BBSRC studentship BB/F01693X/1. SCF and DFC received support from NIH Grant 1R43RR033149-01A1. We are grateful to the skills and attention of Recombinetics staff, The Roslin Institute farm staff led by David Chisholm and Dr. Charles Looney of Texas A&M for collection of Nelore oocytes by OPU.

References

Boulanger L, Pannetier M, Gall L, Allais-Bonnet A, Elzaiat M, Le Bourhis D, Daniel N, Richard C, Cotinot C, Ghyselinck NB, Pailhoux E (2014) FOXL2 is a female sex-determining gene in the goat. Curr Biol 24:404–408

Carbery ID, Ji D, Harrington A, Brown V, Weinstein EJ, Liaw L, Cui X (2010) Targeted genome modification in mice using zinc-finger nucleases. Genetics 186:451–459

Carlson DF, Tan W, Lillico SG, Stverakova D, Proudfoot C, Christian M, Voytas DF, Long CR, Whitelaw CBA, Fahrenkrug SC (2012) Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci 109:17382–17387

Clark AJ, Whitelaw CBA (2003) A future for transgenic livestock. Nat Rev Genet 4:825–833

Clark KJ, Carlson DF, Fahrenkrug SC (2007) Pigs taking wing with transposons and recombinases. Genome Biol 8(Suppl 1):S13

Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibé B, Bouix J, Caiment F, Elsen JM, Eychenne F, Larzul C, Laville E, Meish F, Milenkovic D, Tobin J, Charlier C,

Georges M (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet 38:813–818

Fahrenkrug SC, Blake A, Carlson DF, Doran T, Van Eenennaam A, Faber D, Galli C, Gao Q, Hackett PB, Li N, Maga EA, Muir WM, Murray JD, Shi D, Stotish R, Sullivan E, Taylor JF, Walton M, Wheeler M, Whitelaw B, Glenn BP (2010) Precision genetics for complex objectives in animal agriculture. Anim Sci 88:2530–2539

Geurts AM, Cost GJ, Freyvert Y, Zeitler B, Miller JC, Choi VM, Jenkins SS, Wood A, Cui X, Meng X, Vincent A, Lam S, Michalkiewicz M, Schilling R, Foeckler J, Kalloway S, Weiler H, Ménoret S, Anegon I, Davis GD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Jacob HJ, Buelow R (2009) Knockout rats via embryo microinjection of zinc-finger nucleases. Science 325:433

Grobet L, Martin LJ, Poncelet D, Pirottin D, Brouwers B, Riquet J, Schoeberlein A, Dunner S, Ménissier F, Massabanda J, Fries R, Hanset R, Georges M (1997) A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. Nat Genet 17:71–74

Hauschild J, Petersen B, Santiago Y, Queisser AL, Carnwath JW, Lucas-Hahn A, Zhang L, Meng X, Gregory PD, Schwinzer R, Cost GJ, Niemann H (2011) Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. Proc Natl Acad Sci 108:12013–12017

Ivics Z, Mátés L, Yau TY, Landa V, Zidek V, Bashir S, Hoffmann OI, Hiripi L, Garrels W, Kues WA, Bösze Z, Geurts A, Pravenec M, Rülicke T, Izsvák Z (2014) Germline transgenesis in rodents by pronuclear microinjection of Sleeping Beauty transposons. Nat Protoc 9:773–793

Keele JW, Fahrenkrug SC (2001) Optimum mating systems for the myostatin locus in cattle. J Anim Sci 79:2016–2202

Kurome M, Geistlinger L, Kessler B, Zakhartchenko V, Klymiuk N, Wuensch A, Richter A, Baehr A, Kraehe K, Burkhardt K, Flisikowski K, Flisikowska T, Merkl C, Landmann M, Durkovic M, Tschukes A, Kraner S, Schindelhauer D, Petri T, Kind A, Nagashima H, Schnieke A, Zimmer R, Wolf E (2013) Factors influencing the efficiency of generating genetically engineered pigs by nuclear transfer: multi-factorial analysis of a large data set. BMC Biotechnol 13:43

Leymaster KA, Jenkins T (1993) Comparison of Texel- and Suffolk-sired crossbred lambs for survival, growth, and compositional traits. J Anim Sci 71:859–869

Lillico SG, Proudfoot C, Carlson DF, Stverakova D, Neil C, Blain C, King TJ, Ritchie WA, Tan W, Mileham AJ, McLaren DG, Fahrenkrug SC, Whitelaw CBA (2013) Live pigs produced from genome edited zygotes. Sci Rep 3:2847

Liu X1, Wang Y, Tian Y, Yu Y, Gao M, Hu G, Su F, Pan S, Luo Y, Guo Z, Quan F, Zhang Y (2014) Generation of mastitis resistance in cows by targeting human lysozyme gene to β-casein locus using zinc-finger nucleases. Proc Biol Sci 281:1780

McPherron AC, Lee SJ (1997) Double muscling in cattle due to mutations in the myostatin gene. Proc Natl Acad Sci 94:12457–12461

McPherron AC, Lawler AM, Lee SJ (1997) Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. Nature 387:83–90

Ménissier F (1976) Comments on optimization of cattle breeding schemes: beef breeds for suckling herds. A review. Ann Genet Sel Anim 8:71–87

Ritchie WA, King T, Neil C, Carlisle AJ, Lillico S, McLachlan G, Whitelaw CBA (2008) Transgenic sheep designed for transplantation studies. Mol Reprod Dev 76:61–64

Tan WS, Carlson DF, Walton MW, Fahrenkrug SC, Hackett PB (2012) Precision editing of large animal genomes. Adv Genet 80:37–97

Tan W, Carlson DF, Lancto CA, Garbe JR, Webster DA, Hackett PB, Fahrenkrug SC (2013) Efficient nonmeiotic allele introgression in livestock using custom endonucleases. Proc Natl Acad Sci USA 110:16526–16531

van Wagtendonk-de Leeuw A (2006) Ovum pick up and in vitro production in the bovine after use in several generations: a 2005 status. Theriogenology 65:914–925

Zhang C, Wang L, Ren G, Li Z, Ren C, Zhang T, Xu K, Zhang Z (2014) Targeted disruption of the sheep MSTN gene by engineered zinc-finger nucleases. Mol Biol Rep 41:209–215